UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE FÍSICA

EmoEEG – Recognising people's emotions using electroencephalography

Filipe Miguel Dias Galvão

Mestrado Integrado em Engenharia Biomédica e Biofísica
Perfil em Sinais e Imagens Médicas

Versão Definitiva

Dissertação orientada por:
Manuel João Caneira da Fonseca

2020

Acknowledgments

I'd like to thank Prof. Manuel da Fonseca for being a reliable supervisor, providing me with his expertise and motivation throughout this work. I also thank Soraia Alarcão and João Oliveira for sharing their invaluable knowledge which helped me greatly as well.

On a more personal level, I would like to thank my parents, for all the patience over these long years that I have been studying and for their precious support. I would also like to thank my sister for always cheering me up with her humorous remarks, or when nagging me to play piano.

And finally, I thank my friends for always being there when I needed them, either for sharing silly jokes, playing video and board games, or drinking a much-needed beer.

Resumo

As emoções desempenham um papel fulcral na vida humana, estando envolvidas numa extensa variedade de processos cognitivos, tais como tomada de decisão, perceção, interações sociais e inteligência. As interfaces cérebro-máquina (ICM) são sistemas que convertem os padrões de atividade cerebral de um utilizador em mensagens ou comandos para uma determinada aplicação. Os usos mais comuns desta tecnologia permitem que pessoas com deficiência motora controlem braços mecânicos, cadeiras de rodas ou escrevam. Contudo, também é possível utilizar tecnologias ICM para gerar output sem qualquer controle voluntário. A identificação de estados emocionais é um exemplo desse tipo de feedback. Por sua vez, esta tecnologia pode ter aplicações clínicas tais como a identificação e monitorização de patologias psicológicas, ou aplicações multimédia que facilitem o acesso a músicas ou filmes de acordo com o seu conteúdo afetivo.

O interesse crescente em estabelecer interações emocionais entre máquinas e pessoas, levou à necessidade de encontrar métodos fidedignos de reconhecimento emocional automático. Os autorrelatos podem não ser confiáveis devido à natureza subjetiva das próprias emoções, mas também porque os participantes podem responder de acordo com o que acreditam que os outros responderiam. A fala emocional é uma maneira eficaz de deduzir o estado emocional de uma pessoa, pois muitas características da fala são independentes da semântica ou da cultura. No entanto, a precisão ainda é insuficiente quando comparada com outros métodos, como a análise de expressões faciais ou sinais fisiológicos. Embora o primeiro já tenha sido usado para identificar emoções com sucesso, ele apresenta desvantagens, tais como o fato de muitas expressões faciais serem "forçadas" e o fato de que as leituras só são possíveis quando o rosto do sujeito está dentro de um ângulo muito específico em relação à câmara. Por estes motivos, a recolha de sinais fisiológicos tem sido o método preferencial para o reconhecimento de emoções. O uso do EEG (eletroencefalograma) permite-nos monitorizar as emoções sentidas sob a forma de impulsos elétricos provenientes do cérebro, permitindo assim obter uma ICM para o reconhecimento afetivo.

O principal objetivo deste trabalho foi estudar a combinação de diferentes elementos para identificar estados afetivos, estimando valores de valência e ativação usando sinais de EEG. A análise realizada consistiu na criação de vários modelos de regressão para avaliar como diferentes elementos afetam a precisão na estimativa de valência e ativação. Os referidos elementos foram os métodos de aprendizagem automática, o género do indivíduo, o conceito de assimetria cerebral, os canais de elétrodos utilizados, os algoritmos de extração de

características e as bandas de frequências analisadas. Com esta análise foi possível criarmos o melhor modelo possível, com a combinação de elementos que maximiza a sua precisão.

Para alcançar os nossos objetivos, recorremos a duas bases de dados (AMIGOS e DEAP) contendo sinais de EEG obtidos durante experiências de desencadeamento emocional, juntamente com a autoavaliação realizada pelos respetivos participantes. Nestas experiências, os participantes visionaram excertos de vídeos de conteúdo afetivo, de modo a despoletar emoções sobre eles, e depois classificaram-nas atribuindo o nível de valência e ativação experienciado.

Os sinais EEG obtidos foram divididos em epochs de 4s e de seguida procedeu-se à extração de características através de diferentes algoritmos: o primeiro, segundo e terceiro parâmetros de Hjorth; entropia espectral; energia e entropia de wavelets; energia e entropia de FMI (funções de modos empíricos) obtidas através da transformada de Hilbert-Huang. Estes métodos de processamento de sinal foram escolhidos por já terem gerado resultados bons noutros trabalhos relacionados. Todos estes métodos foram aplicados aos sinais EEG dentro das bandas de frequência alfa, beta e gama, que também produziram bons resultados de acordo com trabalhos já efetuados.

Após a extração de características dos sinais EEG, procedeu-se à criação de diversos modelos de estimação da valência e ativação usando as autoavaliações dos participantes como "verdade fundamental". O primeiro conjunto de modelos criados serviu para aferir quais os melhores métodos de aprendizagem automática a utilizar para os testes vindouros. Após escolher os dois melhores, tentámos verificar as diferenças no processamento emocional entre os sexos, realizando a estimativa em homens e mulheres separadamente. O conjunto de modelos criados a seguir visou testar o conceito da assimetria cerebral, que afirma que a valência emocional está relacionada com diferenças na atividade fisiológica entre os dois hemisférios cerebrais. Para este teste específico, foram consideradas a assimetria diferencial e racional segundo pares de elétrodos homólogos. Depois disso, foram criados modelos de estimação de valência e ativação considerando cada um dos elétrodos individualmente. Ou seja, os modelos seriam gerados com todos os métodos de extração de características, mas com os dados obtidos de um elétrodo apenas. Depois foram criados modelos que visassem comparar cada um dos algoritmos de extração de características utilizados. Os modelos gerados nesta fase incluíram os dados obtidos de todos os elétrodos, já que anteriormente se verificou que não haviam elétrodos significativamente melhores que outros. Por fim, procedeu-se à criação dos modelos com a melhor combinação de elementos possível, otimizaram-se os parâmetros dos mesmos, e procurámos também aferir a sua validação. Realizámos também um processo de classificação emocional associando cada par estimado de valores de valência e ativação ao quadrante correspondente no modelo circumplexo de afeto. Este último passo foi necessário para conseguirmos comparar o nosso trabalho com as soluções existentes, pois a grande maioria

delas apenas identificam o quadrante emocional, não estimando valores para a valência e ativação.

Em suma, os melhores métodos de aprendizagem automática foram RF (*random forest*) e KNN (*k-nearest neighbours*), embora a combinação dos melhores métodos de extração de características fosse diferente para os dois. KNN apresentava melhor precisão considerando todos os métodos de extração menos a entropia espectral, enquanto que RF foi mais preciso considerando apenas o primeiro parâmetro de Hjorth e a energia de wavelets. Os valores dos coeficientes de Pearson obtidos para os melhores modelos otimizados ficaram compreendidos entre 0,8 e 0,9 (sendo 1 o valor máximo).

Não foram registados melhoramentos nos resultados considerando cada género individualmente, pelo que os modelos finais foram criados usando os dados de todos os participantes. É possível que a diminuição da precisão dos modelos criados para cada género seja resultado da menor quantidade de dados envolvidos no processo de treino.

O conceito de assimetria cerebral só foi útil nos modelos criados usando a base de dados DEAP, especialmente para a estimação de valência usando as características extraídas segundo a banda alfa.

Em geral, as nossas abordagens mostraram-se a par ou mesmo superiores a outros trabalhos, obtendo-se valores de acurácia de 86.5% para o melhor modelo de classificação gerado com a base de dados AMIGOS e 86.6% usando a base de dados DEAP.

Palavras-chave: EEG, valência, ativação, extração de características, aprendizagem automática.

Abstract

Emotion recognition is a field within affective computing that is gaining increasing relevance and strives to predict an emotional state using physiological signals. Understanding how these biological factors are expressed according to one's emotions can enhance the human-computer interaction (HCI). This knowledge, can then be used for clinical applications such as the identification and monitoring of psychiatric disorders. It can also be used to provide better access to multimedia content, by assigning affective tags to videos or music.

The goal of this work was to create several models for estimating values of valence and arousal, using features extracted from EEG signals. The different models created were meant to compare how various elements affected the accuracy of the model created. These elements were the machine learning techniques, the gender of the individual, the brain asymmetry concept, the electrode channels, the feature extraction methods and the frequency of the brain waves analysed. The final models contained the best combination of these elements and achieved PCC values over 0.80. As a way to compare our work with previous approaches, we also implemented a classification procedure to find the correspondent quadrant in the valence and arousal space according to the circumplex model of affect. The best accuracies achieved were over 86%, which was on par or even superior to some of the works already done.

Keywords: EEG, valence, arousal, feature extraction, machine learning.

Table of Contents

List of Figures

List of Tables

Capítulo 1

Introduction

In this chapter we outline the motivation that lead to this work by providing a small context regarding emotion recognition via BCI. We also mention the goals we mean to achieve with our work, the proposed solution along with its main contributions, and the structure of this document.

1.1 Motivation

Emotions play an undeniably important role in human lives. They are involved in a plethora of cognitive processes such as decision-making, perception, social interactions and intelligence [1]. Brain computer interfaces (BCI) are systems that translate the brain activity patterns of a user into messages or commands for an interactive application [2]. In 1973, Jacques Vidal coined the term after developing a system that used the visual evoked potentials of the cortex to determine a person's direction of gaze, and use that information to move a cursor on a computer [3]. This is an example of an active BCI, in which a user controls a device using conscious commands that are relayed to an external application. The most common uses for this technology allow for individuals who are motor-impaired to control mechanical arms, wheelchairs, or to write. A passive BCI however, uses brain signals to generate an output without any voluntary control. Emotional states are an example of this kind of feedback.

Due to the recent interest shown in establishing emotional interactions between humans and computers, the identification of a person's emotional state became a need. This can be done through self-reports, which can be unreliable due to the subjective nature of emotions themselves, but also because participants may answer according to what they believe others would answer. Emotional speech is an effective way of deducing one's emotional state, since many speech characteristics are independent of semantics or culture. However, the accuracy still falls short when compared to other methods such as the analysis of facial expressions or physiological signals. Although the former has been used to successfully identify emotions, it suffers from drawbacks like the fact that many facial expressions are "unnatural" and the fact

that the readings are only possible when the face of the subject is within a very specific angle towards the camera. As such, the collection of physiological signals, has been the preferred method for emotion recognition. The use of EEG (electroencephalogram) allows us to monitor the emotions felt as electric impulses coming from the brain, thus allowing us to obtain a passive BCI for affective recognition.

In order to predict an individual's emotion or emotional state accurately, one must extract meaningful features from the physiological signals, and label them accordingly. After that, a model can be trained using machine learning algorithms to assess a person's affective state using new found physiological signals. To generate a model, databases containing self-assessed values of valence and arousal, taken from individuals when experiencing emotional content are used. The most common approach is to guess which quadrant within the circumplex model of affect the emotional state of an individual is. However, inserting the physiological signals into classes tells us the emotion but not the degree to which it is being expressed. Taking this into consideration, a regression analysis will be performed, in order to estimate individual values of valence and arousal of an individual.

This work was conducted at LASIGE, a research unit at the Department of Informatics, Faculty of Sciences, University of Lisbon, in the context of the project Awareness While Experiencing and Surfing On Movies through Emotions (AWESOME), supported by the Fundação para a Ciência e Tecnologia (FCT) under LASIGE Strategic Project - UID/CEC/00408/2019, and under project AWESOME - PTDC/CCI- INF/29234/201.

1.2 Goals

The main objective of this work was to study the combination of different elements for estimating valence and arousal values using EEG signals. The analysis performed consisted in creating various regression models in order to assess how these different elements affect the accuracy in valence and arousal estimation. Said elements were the machine learning methods, the gender of the individual, the brain asymmetry concept, the electrode channels used, the feature extraction algorithms and the frequency bands analysed. After this analysis, we have gained an understanding that allowed us to create the best model for affective estimation.

Hopefully this work will provide insight into what procedures to take in order to predict an individual's emotion or emotional state as accurately as possible.

1.3 Developed Solution

To accomplish our goals, we resorted to using two datasets (AMIGOS and DEAP) containing EEG signals taken during emotion elicitation experiments, along with the self-assessment performed by the respective participants. Several methods of feature extraction were performed on each epoch of the EEG signals, considering alpha, beta and gamma frequencies, namely the three Hjorth parameters (activity, mobility and complexity), spectral entropy, wavelet energy and entropy and IMF energy and entropy.

After that, several regression analyses were performed in order to access the best regression techniques for estimating valence and arousal values, using the participants' self-assessment as ground truth as well as all the extracted features. After choosing the two best regressors, we tried to verify any differences in emotion processing between genders, by performing the estimation on men and women separately. The brain asymmetry concept, which states that emotional valence is related with differences in physiological activity between the two brain hemispheres, was also tested.

After that, each electrode was tested individually for valence/arousal estimation using all the extracted features. After noticing that there were no electrodes that were better than others regarding estimation performance, each feature extraction method was tested individually, considering all the electrodes.

Finally, after finding the best model for estimation, we evaluated the emotional state of each participant by associating each estimated pair of valence/arousal values to the corresponding quadrant in the circumplex model of affect, thus discovering the participants' emotional state. This was done mainly to compare our work with previous approaches.

1.4 Contributions

Most works that deal with emotion recognition by analysing physiological signals, do so by attempting to directly find the correct class using machine learning classification. Contrary to those, in our work, the final classification phase is preceded by a regression phase. This way, not only can we know the emotional state of a person, but also, to what degree that emotional state is actually being manifested. We achieve that by predicting the valence and arousal values via regression.

The wide array of tests performed also allows for the comparison of several different elements, in order to find out how each affect the overall emotion recognition task. As such, the main contribution of this work stems from these tests, from which we were able to assess which were the best machine learning algorithms for emotion recognition, test the validity of the brain asymmetry concept, and discover which were the best feature extraction methods.

1.5 Document Structure

Chapter 2 focuses on most of the theoretical aspects of emotions, mainly what they are, how they can be represented, and how they are expressed in terms of physiological signals, including EEG. Chapter 3 mentions some of the works already performed in emotion recognition via EEG. It describes their approaches and briefly states their results. Chapter 4 describes this work's approach to emotion classification. It outlines the used datasets, algorithms and regression analyses which lead to the final solution. In chapter 5, we present the evaluation of the aforementioned procedure. Chapter 6 presents a discussion of the results obtained in the previous chapter. And finally, chapter 7 outlines the conclusions and future work.

Capítulo 2

Background Theory

In this chapter we attempt to enlighten the reader on a few notions concerning emotion classification using EEG. First, we explain what emotions are and how they can be represented. Then we introduce the concept of affective computing and how physiological signals, such as EEG can be used to recognize a subject's emotional state. Finally, we expose a few ideas regarding EEG, which was the technique considered in our effort to identify people's emotional state. This chapter is meant to give context only, as the actual description of the approaches taken to our specific problem can be read in the next chapter.

2.1 Emotions

Emotions are generated whenever a perception of an important change in the environment or in the physical body appear. There are currently two main scientific ways of explaining the nature of emotions. According to the cognitive appraisal theory, emotions are judgments about the extent that the current situation meets our goals or favours our personal well-being [4] [5]. The fundamental proposition is that the evaluation of a given situation causes the emotional response in accordance with a set of psychobiological laws, established by our knowledge and beliefs [5]. How these representations are appraised with respect to their significance for personal benefit is what determines the emotional state, and they can vary according to one's personality [5]. Alternatively, William James and Carl Lange have argued that emotions are perceptions of changes in our body such as heart rate, breathing rate, perspiration, and hormone levels [4]. Each emotion is therefore followed by a specific pattern of physiological arousal. The James-Lange theory was one of the first theories of emotion in modern psychology and has been modified several times since its conception. Today, the main premise still stands, although scholars now state that there are other factors that modulate emotional experience besides physiological feedback. Either way, emotions are conscious experiences characterized by intense mental activity and a certain degree of pleasure or displeasure.

Emotion classification is the means by which different sets of emotions can be distinguished. There are essentially two fundamental approaches for this task: one can consider emotions as discrete and fundamentally different constructs, or that emotions can be characterized on a dimensional basis in groupings.

In discrete emotion theory, all humans are thought to have an innate set of basic emotions that are cross-culturally recognizable. These basic emotions are described as "discrete" because they are believed to be distinguishable by an individual's facial expression and biological processes. Theorists have conducted studies to determine which emotions are basic. A popular example is Paul Ekman and his colleagues' cross-cultural study of 1992, in which they concluded that the six basic emotions are anger, disgust, fear, happiness, sadness, and surprise [6]. Ekman explains that there are particular characteristics attached to each of these emotions, allowing them to be expressed in varying degrees. Each emotion acts as a discrete category rather than an individual emotional state.

One of the most common frameworks in the emotions field proposes that affective experiences are best characterized by two main dimensions: arousal and valence. Valence, as used in psychology, especially in discussing emotions, means the intrinsic attractiveness or averseness of an event, object, or situation. However, the term is also used to characterize and categorize specific emotions. Arousal is the physiological and psychological state of being awoken or of sense organs stimulated to a point of perception. Emotions can be mapped out on a chart (Figure 2.1) modelling the range of arousal (high to low) and valence (pleasure to displeasure) that is experienced during a particular emotion. For example, in the top right corner are the emotions with high arousal and high valence, which include excited, astonished, delighted, happy, and pleased. These emotions are all examples of positive emotions that are high in arousal and high in valence. In the opposite corner is the low valence and low arousal section, containing miserable, depressed, bored, and tired as some examples.

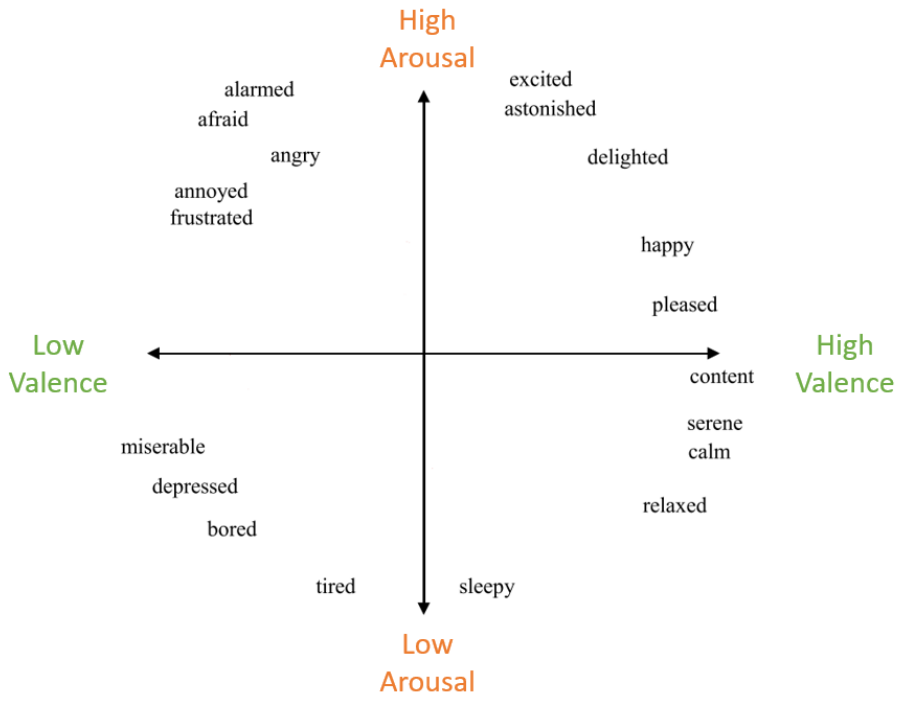

Figure 2.1- Valence-arousal circumplex chart

2.2 Physiological Signals and Feature Extraction

Emotions play an undeniably important role in human lives. They can be expressed either by speech, body language or facial expressions. The recent decades have seen a rise in different forms of human-computer interaction (HCI), either for entertainment, communication, or to overcome disabilities. Affective computing is the study and development of systems that can perceive, identify or process human emotions. It could also serve as the means to give machines the ability to act emotionally [7].

Emotion recognition from humans starts with collecting emotional information via sensors which capture data about the user's physical state or behaviour. The conventional methods basically utilize audio and visual attributes to model human emotional responses, such as speech, facial expressions, and body gestures. More recently, accessing physiological responses has gained increasing attention in characterizing the emotional states. These methods include the measurement of heart rate (HR), galvanic skin response (GSR), electromyography (EMG), electrocardiography (ECG) and respiration rate (RR). Measurements over brain activity can also be recorded with the use of functional magnetic resonance imaging (fMRI), positron emission tomography (PET) or electroencephalography (EEG), the latter being the main focus of this work. It is also possible to combine different modalities such as EEG and ECG. Although EEG offers poor spatial resolution and requires the careful positioning of several electrodes in the user’s head, it provides great time resolution, fast data acquisition, is non-invasive, and is

inexpensive. Furthermore, the use of portable and wireless EEG devices (e.g. Emotiv) can help mitigate the setup inconvenience.

During an EEG emotion recognition experiment three important steps must be taken. The first refers to the emotion elicitation, which must be as efficient as possible. Emotions are usually triggered through the use of videos, music or images, previously chosen for this effect. Second and third steps have to do with the captured data pre-processing and classification, respectively. Their effectiveness, however, is highly dependent on the emotion elicitation. If the subjects have not effectively become emotionally aroused during the emotion elicitation steps, the respective signals would not hold the corresponding emotional information, resulting in an inaccurate emotion classification process.

Recognizing emotional information requires the extraction of meaningful patterns from the gathered data. This is accomplished using a wide variety of feature extraction methods that process different modalities, which are later fed to a classifier in the form of a feature vector. When referring to EEG emotion recognition, feature vectors can be comprised of statistical features (mean, standard deviation, kurtosis, etc), spectrum analysis (power spectrum densities via Fourier transform, empirical mode decomposition or wavelet transform, etc), or digital signal analysis as a whole (ARMA[1] models, high order crossing, etc). It is also common to extract features within dipoles of electrodes from opposite hemispheres in accordance with hypothesis relating brain asymmetry with emotional expression [8] [9].

After extracting the features from the signals and assigning their respective "emotional labels" (obtained through self-assessment by participants), we have all the necessary components to train a model for emotion classification via machine learning.

2.3 Electroencephalography (EEG)

The firing of neurons in the brain trigger voltage changes. The electrical activity measured by the electrodes in an EEG headset corresponds to the field potentials resulting from the combined activity of many individual neuronal cells in the brain cortex.

The frequency of EEG measurements ranges from 1 to 80Hz, with amplitudes of 10 to 100 microvolts [4]. Signal frequencies are divided into different bands, since specific frequency waves are normally more prominent in particular states of mind, namely the delta (1-4 Hz), theta (4-7 Hz), alpha (8-13 Hz), beta (13-30Hz), and gamma (>30 Hz) bands (see Figure 2.2). Delta are usually the lowest in frequency and slowest waves, and are also prominent during NREM sleep. Theta waves are associated with subconscious activities, such as dreaming, and

[1] Auto regressive and moving average.

are present in meditative states of mind. Alpha waves predominantly originate during wakeful relaxation mental states with the eyes closed, and are most visible over the parietal and occipital lobes [10]. Intense alpha wave activity has also been correlated to brain inactivation. Beta wave activity, on the other hand, is related to an active state of mind, most prominent in the frontal cortex during intense focused mental activity [10]. Lastly, gamma rhythms are thought to be associated with intense brain activity for the purpose of running certain cognitive and motor functions.

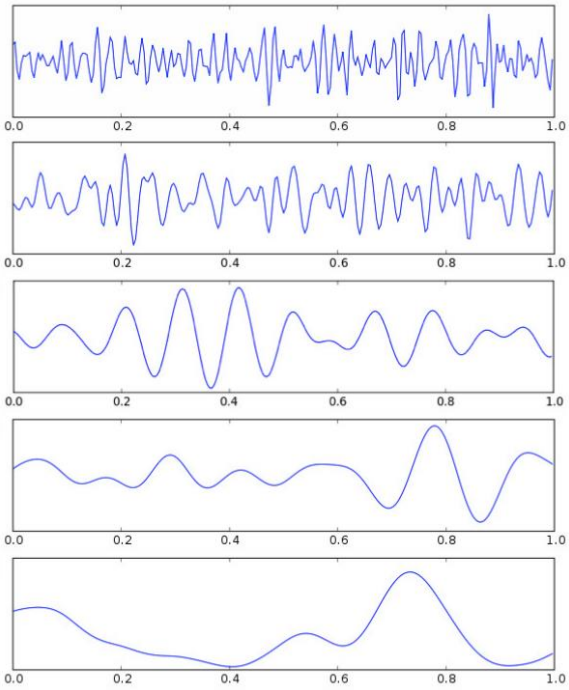

Figure 2.2- The five main brain waves. From bottom to top: delta, theta, alpha, beta and gamma, respectively.

The International 10–20 system is a standardized method to describe and apply the location of scalp electrodes in the context of an EEG exam. The system is based on the relationship between the location of an electrode and the underlying area of the brain, specifically the cerebral cortex. The numbers "10" and "20" refer to the distances between adjacent electrodes, which are either 10% or 20% of the total front–back or right–left distance of the skull.

Each electrode site is represented by a letter identifying the lobe, and a number to identify the hemisphere location (Figure 2.3). “F” stands for Frontal, “T” for Temporal, “C” for Central, “P” for Parietal, and “O” for Occipital. “z” (zero) refer to electrodes placed along the sagittal plane. Even numbers refer to electrode positions on the right hemisphere, while odd numbers refer to the left one. Four anatomical landmarks are used for the correct positioning of the electrodes: nasion (the point between the forehead and nose), inion (the lowest point of the skull from the back of the head, indicated by a prominent bump), and the pre-auricular points anterior to the ear. Occasionally, one might use A1 and A2 electrodes. These refer to the prominent bone process found just behind the outer ear, and are particularly used in polysomnography.

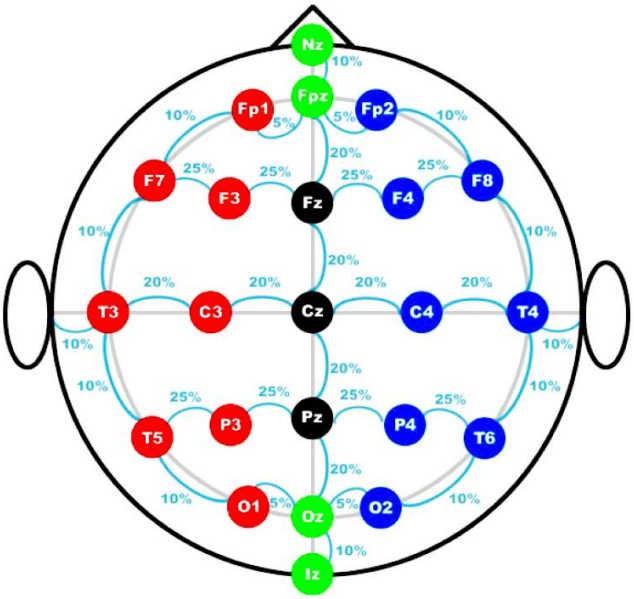

Figure 2.3- Electrode channels arrangement according to the 10-20 system

Capítulo 3

State of the Art

Before moving on to the main focus of our work, it is useful to point out some of the most important aspects of the attempts already made by researchers in the field of emotional recognition and classification. The next sections divide the work according to the different approaches taken and how successful they were. As stated before, EEG emotion recognition experiments are usually carried out in three steps. The first step is concerned with the emotion elicitation task, which is accomplished either through watching pictures, videos, listening to music or attempting to recall emotional experiences. The second step has to do with the extraction of features from the EEG signal which best represent the emotional content. There are several different ways of doing so, and all of them have different levels of complexity and diverse, whether or not there is a clear paradigm. The final step has to do with the classification of these features, which is heavily reliant on how well said emotion was expressed, as well as the classification method. Naturally, if the features extracted are not reliable in the first place, sorting the emotional content will be at fall, and the classification task will ultimately fail.

3.1 Databases and Emotion Elicitation

Only a small number of benchmark emotional EEG databases with categorized emotions are publicly available for use and to test a new system. As mentioned before there are several different ways of making people experience emotions. Music videos are a popular approach, and the Database for Emotion analysis using Physiological signals (DEAP) [11] is likely the most used. This database is made of a collection of several physiological signals (EEG, EMG, EOG, GSR, face recordings, among others) of 32 people (16 men and 16 women) who watched 40 music videos, previously chosen through an online enquiry. The videos were assigned a score of valence, arousal, liking and dominance through a self-assessment manikin (SAMs) (Figure 3.1). The DEAP database is one of the most used in research on emotion recognition ([12] [13] [14], among others). It is vast, containing physiological signals that do not only include EEG. It is also made of data taken from both men and women on equal number. This is

also a great advantage since men and women have shown perceive emotions differently [15] and having an equal number should result in a less biased classifier.

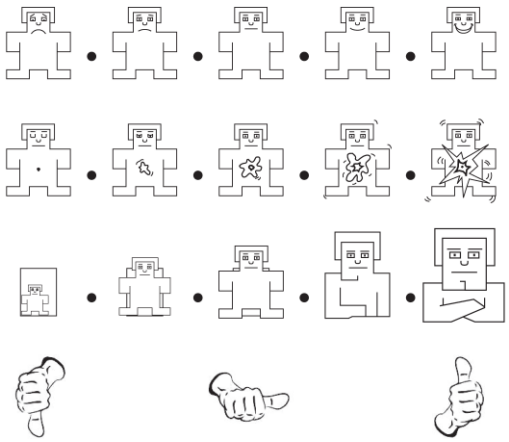

Figure 3.1- Self-Assessment Manikin (SAM). Each line represents a scale for valence, arousal, dominance and liking.

Another database of the same type, which is accessible online is the Multimodal Database for Affect Recognition and Implicit Tagging (MANHOB) [16]. It is comprised of EEG recordings, peripheral physiological signals, eye tracking data and face video recordings of subjects watching 20 film clips. The data was synchronized, and also taken from a large group of people (11 men and 16 women).

The SEED database contains EEG signals of 15 subjects, recorded while they were watching emotional film clips. For the feedback, participants were told to report their emotional reactions to each film clip by completing a questionnaire immediately after watching each clip. Facial videos and EEG data were recorded simultaneously [17].

AMIGOS is a recently published dataset that contains EEG and ECG signals taken from 40 individuals whilst they watched 16 film clips with varying duration [18]. The subjects also performed a self-assessment and the data was subsequently used, not only for emotion detection and classification, but also for personality inference.

Other popular approaches involve the use of pictures and sounds, which were chosen specifically to trigger emotional states. The International Affective Picture System (IAPS) has been successfully used for emotional recognition a number of times [19] [20], as well as the International Affective Digitized Sound System (IADS) [21]. However, these datasets do not include any physiological signals whatsoever.

Of course, the main advantage of using online databases is that we do not need to perform the emotion elicitation part, and also has access to the ground truth present in the self-ratings. Still, some people have chosen to produce their own databases and have successfully managed to identify a series of emotions using Hollywood Oscar film soundtracks [22] and Oscar film

clips [23]. Some have also attempted to generate emotional states by telling participants to recall personal emotional experiences [24].

How effectively the desired emotion is inducted is a significant query that is relatively difficult to answer. Aspects like personality, personal experiences, and a particular subject's mood at the time that the experiment is conducted dramatically influence the way someone reacts emotionally in view of that kind of image. As a result, as soon as the emotion is induced by a stimulus outside the efforts of the subject, how effectively the emotion is actually evoked is questioned, and consequently, how representative the EEG activity is with regard to a particular emotion. Taking this into consideration Petratonakis, et al [25] decided to use the Pictures of Facial Affect database (POFA), which comprises of peoples' facial expressions when experiencing certain emotions. This is supported by the Mirror Neuron System (MNS) theory, which states that when individuals observe an action done by another individual, they have the same or akin brain activity as if they did the corresponding action themselves.

Although the affective information from image, video, and audio stimuli has been extensively studied, olfactory stimuli [26], written words [27], food stimuli (enriched by emotional stimuli) [28], and games [29] have also been used as elicitation methods in a number of studies as ways to assess human emotional state by investigating physiological signals.

3.2 Feature Extraction and Selection

When it comes to any task involving EEG, the first thing to decide is the number of electrodes to be used. In the interest of emotion recognition, they can vary from only 3 electrodes [30] to a total of 64 electrodes [23] [31]. The usual number of electrodes revolves around 32 [11] [13] [22] [32] [33] (according to the 10-20 system). Naturally when considering the number of electrodes, time interval required to set up an EEG device, comfort level of subjects, system usability, and number of features to be processed, it is logical, from this standpoint, that fewer electrodes should be utilized. However, an approach that is also quite common is to collect EEG signals from many electrodes and posteriorly decrease the number after processing the data using statistical techniques meant for dimensionality reduction. It is also worth noting that not only the monopoles are taken into account, but also dipoles, which are meant to describe the asymmetry between the hemispheres ([22] [25] [30], are only a few). This concept has been used in many experiments, and states that the difference in the activity between the hemispheres reflects the emotional positivity valence. A higher left activity is related with a positive emotion (high valence), whereas a high right activity is related with a negative emotion (low valence). When the number of electrodes sits in 3-32, the ones usually used are the ones belonging to the frontal and parietal lobes since those have produced the best results, among the most studies.

Regarding the brain waves, most people prefer to use the set comprised of theta, alpha, beta and gamma. Some use delta as well [20] [22] [23] or even the full EEG without separating the frequencies [30] [34]. In [19] [25], Petratonakis et. al decide to use only alpha and beta frequencies, since those had generated the best results in previous work. The same for Zhang et. al [35] [36] who decided to use only beta frequencies. As stated before, it is also possible to select the best set of frequencies, based on dimensionality reduction.

Once the electrodes have been chosen, along with the frequencies to use, we must decide upon the feature extraction algorithm that best describes the emotions soon to identify. There are several different methods, and it is quite common for certain authors to choose more than one. The most used methods are the Fourier Transform such as the Short-time Fourier Transform (STFT) or Discrete Fourier Transform (DFT) [12] [17], statistical (Mean, Standard deviation, Kurtosis, Skewness, Correlation) [20] [37], Hjorth Parameters (HP) [13] [34] [37], Power Spectral Density (PSD) [11] [22] [23] [32] (among others), Wavelet Transform (WT) [14] [23], Empirical Mode Decomposition (EMD)/Hilbert-Huang Spectrum (HHS) [12] [25] [35] (among others), Entropy such as the Differential Entropy (DE) [17] [31], Approximate Entropy (AE) [23], Sample Entropy (SE) [35], or Wavelet Entropy (WE) [14] [23], Higher Order Crossings (HOC) [19], Common Spatial Patterns (CSP) [24] [38], Auto Regressive Models (AR) [36], Fractal Dimensions (mainly the Higuchi Fractal Dimension (HFD)) [13] [23] [30]. As stated before, it is possible to consider measures taken from individual electrodes (monopoles) or pares of electrodes (dipoles). In the second case, the features are usually described as asymmetry measures. These include Differential and Rational Power Spectral Asymmetry (DPSA and RPSA) [11] [22] [32] (among others), Mutual Entropy (ME)/Mutual Information (MI) [37], Asymmetric Correlation (ACorr) [19], or Asymmetric Coherence (AC) [37].

The process of feature extraction usually generates large amounts of data. In order to ease the computational burden of the next stage (emotion classification), it is common practice to select the best features, and once again, this can be achieved through several ways. Some of the most common are the Principal Component Analysis (PCA) [23] [31] [32], Minimum Redundancy Maximum Relevance (mRMR) [12] [13] [31], Fisher Linear Discriminant (FLD)/ Linear Discriminant Analysis [11] [23] [37]. In case of EMD usage, it is also common to simply choose the intrinsic mode functions (IMFs) which contribute for the most variance [35] [36]. A table containing both the feature extraction and selection methods used in EEG emotion recognition works can be consulted in Appendix A (Table A1).

3.3 Emotion Classification

When it comes to recognising emotions, practically all authors choose to resort to machine learning classifiers. In most of the works Support Vector Machines (SVM) [19] [22] [23] [25] (among others) was used followed by K-Nearest Neighbours (KNN) [15] [17] [34] (among others). Other machine learning classifiers used were Quadratic Discriminant Analysis (QDA) [25], Naive Bayes (NB) [11], Multi-Layer Perceptron (MLP) [22] [37], Artificial Neural Networks (ANN) [34], Deep Learning Networks (DLN) [32], Deep Belief Networks (DBN) [17], and Random Forest (RF) [12]. Although technically not considered a classifier, Logistic Regression (LR) was also used a number of times [17] [31]. It is also usual practice to use more than one machine learning algorithm and then decide which one performs best given the feature vector.

Tackling of the emotion classification problem can be done in one of two ways: we can focus on identifying discrete emotions such as happiness, scared, or disgust [21] [22] [25] (among others); or we can simply focus on finding the quadrant, in the valence/arousal space (see introduction) [11] [19] [37] (among others). In the latter approach authors choose to create two classifiers, one to discern between high/low valence, and the other for high/low arousal. In terms of accuracy, this approach tends to generate better results, since we are not interested in the emotions themselves, but rather a range of arousal/valence values, which can encompass more than one emotion. Authors have also considered including all positive emotions in one class and all negative emotions in another, leaving only two classes to work with, or three, depending on whether one considers the "neutral" emotion or not [17] [23]. This is done by turning valence/arousal values into emotions, using a set of empirical rules described in [11]. A table containing the classes, classification method, accuracy of the various consulted works, as well as the number of individuals present in the study can be seen in Appendix A (Table A2). It also discloses whether the given studies considered any differences in gender regarding the emotion recognition tasks.

3.4 Summary

Although there is little consensus about the best approach to take for the task of identifying emotions via EEG, after some research we can identify a few practices that seem to generate the best results.

There is little consensus about the number of electrodes or which ones are best to use, although the frontal ones seem to produce better results.

Alpha, beta, and gamma waves also appear to be strongly related with emotional expression. Many authors choose these based on past works, without even going through the feature selection part. Looking at the accuracies we can notice that statistical features seem to be

the least reliable for describing emotions. Complex classifiers such as DLN, DBN or MLP are not necessarily the best, since there are works where simpler classifiers like SVM or KNN can generate better results. This leads one to believe that the main aspect affecting the accuracy could be the feature extraction process. Spectral analysis is the most used technique, but PSD is not necessarily the best approach. Applying the Fourier transform to an EEG signal still gives us a somewhat "chaotic" representation. Relatively simple machine learning methods such as SVM tend to be sensitive to noisy data, which means more complex classifiers will be needed. Thus, a good choice of features should be one that produces smoother curves, such as WT or EMD. It is also worth mentioning that entropy-based features like SA, AE and WE achieve results with high accuracy as well.

As for the set of emotions to be recognized, there is little agreement. There are as many works intending to identify basic emotions (or subsets of them), as those focusing on the valence and arousal levels. These last ones appear to show better accuracies. After all, the valence/arousal chart encompasses several emotions into four quadrants, making the classification task much simpler. When the number of emotions to be recognized increases, the accuracy tends to diminish. It is also possible to achieve precise values of valence/arousal instead of ranges of values, but to achieve that one would need to use a regressor. To our knowledge, only [33] and [58] have taken this approach.

As a final note, whether the databases possess similar or equal number of men and women, it is important to add that the physiological (in this case, via EEG) expression of emotions is different according to gender. Among the works consulted, this premise was only taken into account by Lee, et al [45] and Chen, Jing, et al in [15].

Capítulo 4

Experimental Setup

Here we mention the various steps taken in our effort to develop a system for deducing one's emotional state. We begin by describing in detail the datasets used, AMIGOS [18] and DEAP [11]. Then we recount the experiment protocol, which is divided into three steps. The first is the pre-processing step, where the EEG signals were filtered of artefacts. This step was performed using adaptive filtering, and only on the AMIGOS dataset, since we used the already pre-processed signals of the DEAP dataset. The second step is the feature extraction process, where specific algorithms were applied to the EEG signals in order to extract its meaningful characteristics. These algorithms have all been performed in similar experiments and are explained in detail. The last step concerns the regression analysis where machine learning methods are used in order to estimate values of valence and arousal. Later on, these values are used for classification, by connecting ranges of valence/arousal values to a given quadrant of the circumplex model, for comparison with the state of the art.

4.1 Datasets

As stated before, the task of emotion elicitation can be performed using various forms of stimuli. For this experiment we resorted to using videos as the method for triggering the emotions. However, in order to bypass the hassle of acquiring the data, we used two publicly available online datasets: AMIGOS and DEAP.

4.1.1 AMIGOS

This dataset was created with two purposes. Not only did the authors attempt to study people's affective response through the use of physiological signals, but their personality as well. This was done in two different scenarios: one where each participant was alone, and the other where they were part of an audience. At the same time, people's emotions were being elicited by two types of content. The first type consists of short emotional videos

(duration<250s) selected to elicit specific affective states [18]. The second type consists of long videos (duration>14min), that present situations that could elicit various affective states over their duration and where the story and the narrative could give context to the affective responses [18].For the purpose of this work, we concern ourselves only with the scenario in which the participants were alone and watching the short videos, and therefore, only these will be described.

These videos were chosen from two other datasets MANHOB [15] and DECAF [73] so there would be an equal number of videos classified as HVHA, HVLA, LVHA and LVLA (H, L, A and V stand for high, low, arousal and valence respectively), and as further from the origin of the scale as possible. The total number of selected short videos is 16, 4 for each quadrant of the VA space (Table 4.1).

Category	Excerpt's source
HAHV	Airplane, When Mary Met Sally, Hot Shots, Love Actually
LAHV	August Rush, Love Actually, House of Flying Daggers, Mr Bean's Holiday
LALV	Exorcist, My Girl, My Bodyguard, The Thin Red Line
HALV	Silent Hill, Prestige, Pink Flamingos, Black Swan

Table 4.1- The short videos listed with their source (adapted from [11]).

Three types of physiological signals were recorded: EEG, GSR and ECG. The first was the main focus of this work and was recorded using the Emotiv EPOC Neuroheadset, which mitigates the incumbrance of conventional setups given its portability. This arrangement allows the recording of the electrical activity in the scalp using 14 electrode channels: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4 (according to the 10-20 system). The device recorded the signals at a 128 Hz rate, 14-bit resolution. Although the ECG signals, were not used for recognising effective states in our work, they were used during the pre-processing step for noise removal, and were recorded using the Shimmer 2R platform at 256 Hz rate and 12-bit resolution, using three electrodes, two of them placed at the right and left arm crooks and the third one at the internal face of the left ankle as reference [18].

Recordings were performed in a laboratory environment with controlled illumination [18]. Forty healthy participants (13 female and 27 male), aged between 21 and 40 (mean age=28,3), took part in the experiment. Each signed a consent form and read the instructions concerning the experiment. After placing the sensors, and before watching any videos, the participants performed a self-assessment of their arousal, valence and dominance and selected a basic emotion (Neutral, Happiness, Sadness, Surprise, Fear, Anger and Disgust) they felt before the stimulus. Next, 16 videos were presented in a random order in 16 trials, each consisting of: a 5 second baseline recording showing a fixation cross; the display of a small video. Then, at the end of each trial, participants performed a self-assessment of the same dimension as the initial self-assessment, plus the liking, familiarity and emotion that described what they felt during

each video. The self-assessment form used for the short videos experiment can be seen in Figure 4.1. It is similar to the SAM shown in chapter 3 with the addition of the liking, familiarity and emotion assessment.

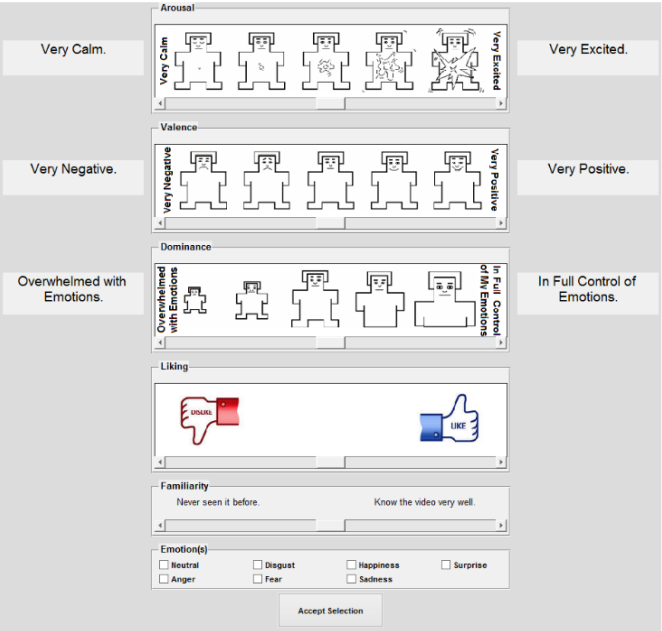

Figure 4.1- Layout of the SAM used for assessing arousal, valence, liking familiarity and basic emotions (taken from [18]).

After the 16 trials, the recording session ended. In total, for the short videos experiment 17 annotations were obtained from each participant (1 at the beginning of the experiment and 1 after each of the 16 short videos) [18].

4.1.2 DEAP

Unlike the AMIGOS, this dataset was built using music videos instead of film clips. The videos were chosen from an initial set of 120 videos, picked from the Last.fm website, both manually and using an algorithm capable of analysing the tags associated with the videos. From there, several one-minute segments were extracted and, using a machine learning approach, the segment that contained the most emotional content was selected. This selection was based on an estimation of valence and arousal through video content-based features such as lighting, colour variance, motion, and loudness. Finally, the 120 one-minute videos were reduced to a set of 40 by using a web-based subjective emotion assessment interface, where volunteers would rate the videos on a scale of 1-9 on valence, arousal and dominance, similarly to a SAM. Then, for each quadrant in the normalized valence/arousal space, the 10 videos that lied closest to the extreme corner of the quadrant were selected. Figure 4.2 shows the score for the ratings of each video and the selected videos highlighted in green.

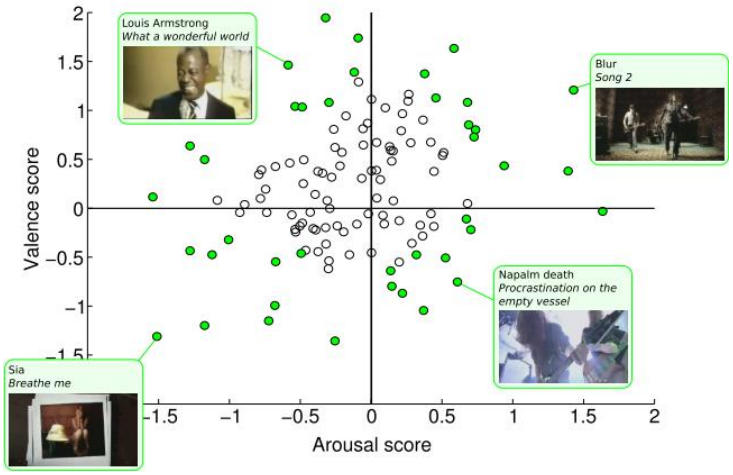

Figure 4.2- Normalized values for the ratings of each video in the online assessment. Videos selected for use in the experiment are highlighted in green (taken from [11]).

EEG was recorded at a sampling rate of 512 Hz using 32 active AgCl electrodes (placed according to the international 10-20 system) using a Biosemi ActiveTwo system. Other physiological signals recorded were GSR, blood volume pressure, respiration pattern, skin temperature, EOG, and EMG. Once again, for the purposes of this work, only EEG is considered.

Thirty-two healthy participants (50% female), aged between 19 and 37 (mean age 26.9), participated in the experiment. They also signed a consent form and were given instructions regarding the experiment. Then the participants performed a practice trial to familiarize themselves with the system. In this unrecorded trial, a short video was shown, followed by a self-assessment by the participant. The real experiment started with a 2-minute baseline recording, during which a fixation cross was displayed to the participant (who was asked to relax during this period). Then the 40 videos were presented in 40 trials, each consisting of the following steps: a 2 second screen displaying the current trial number to inform the participants of their progress; a 5 second baseline recording (fixation cross); the 1-minute display of the music video; self-assessment for arousal, valence, liking and dominance.

After 20 trials, the participants took a short break, during which, they were offered some cookies and non-caffeinated, non-alcoholic beverages. The experimenter then checked the quality of the signals and the electrodes placement and the participants were asked to continue the second half of the test.

4.2 Pre-Processing

The EEG signals obtained through the processes described above are still raw, meaning they still have not been subjected to artefact removal, and as such, still require processing. The techniques applied to the EEG data from both AMIGOS and DEAP were similar, and are essentially a common average reference followed by the application of a blind source separation technique (BSS). However, in our work we chose to perform our own approach to pre-processing, based on adaptive filtering and wavelet thresholding. Contrary to BSS which usually requires further manual input by the user, our approach is entirely automatic. To give us a comparison standpoint, we decided to apply this pre-processing step only to the AMIGOS dataset.

The first step was detrending the signal and eliminating the 50 Hz power line frequency, present in every European power grid, by applying a notch filter. The rest of the filtering process is a bit more complex and will be explained in detail below. This approach requires the use of adaptive filtering techniques.

An adaptive filter is a system with a linear filter that has a transfer function controlled by variable parameters and a means to adjust those parameters according to an optimization algorithm [74]. Let s[n] denote the observed signal which is a combination of the original EEG, x[n] and additive artefact r[n]. Then, if the artefact source v[n] is available from a dedicated channel (e.g. EOG or ECG); an adaptive algorithm (e.g. LMS, RLS, etc.) can be used to derive an artefact-free EEG, x'[n] given that the desired EEG and artefact signal are independent or uncorrelated [74]. An illustration of the use of adaptive filter for artefact removal is shown in Figure 4.3.

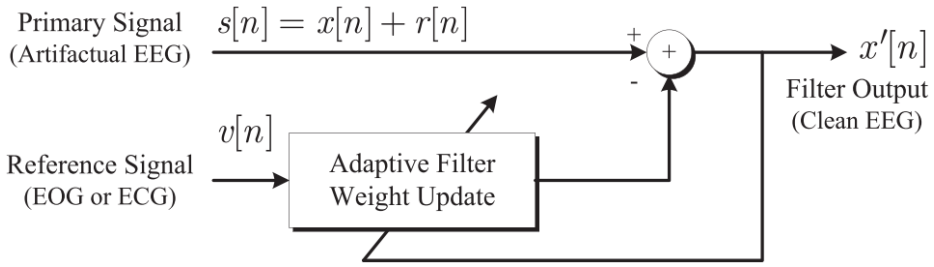

Figure 4.3- Box diagram of an adaptive filter. (taken from [74])

The LMS (least mean squares) algorithm was used in this work to remove the ECG component of the EEG signals, and it works by minimizing the error between de desired signal and the actual signal. The LMS algorithm is fairly simple, fast and effective, making it a good choice for noise removal. Before applying the filtering, the ECG was downsampled from 256 Hz to the same sampling frequency of the EEG which was 128 Hz. The main filter parameters are L and μ, which are respectively the filter order and the rate of adaptation. The optimal values were chosen through a process of trial and error with a few signals. First, different values

of L were tested (L=8, L=16, L=32) for a fixed value of $\mu=10^{-5}$. After obtaining the optimum value of L, several values of μ were tested, multiples of the power of 10 ($10^{-5}10^{-6}$, 10^{-7}, 10^{-8}, etc). If μ is too big, the filter becomes unstable, and if it is too small, the adaptation may turn out too slow [75].

The ECG component typically carries information within the beta and gamma spectrum of an EEG. For alpha frequencies the most prevalent artefacts are EOG. These were removed using wavelet thresholding. Using the discrete wavelet transform (see feature extraction section) one can decompose a signal into a set of detailed and approximate coefficients belonging to different decomposition levels. The A2 level (approximate coefficients belonging to the second level) contains the frequencies alpha, theta and delta. Since EOG artefacts are characterized by high amplitude spikes, it is possible to detect them with the right threshold and the right mother wavelet [76]. After that, all we need to do is to apply the inverse wavelet transform to obtain the signal, free of EOG artefacts. The mother wavelet chosen was db7, based on the results of [76].

Muscle activity artefacts are especially difficult to remove using automatic methods, because of their wide spectrum range and variance in terms of amplitude. However, they should be mitigated to a certain degree since participants were told not to move during the experiment.

Finally, the first 5s of the EEG signals were removed (baseline) and then each signal was averaged to the common reference. The baseline was also removed from the DEAP signals, in which case, it was a 3s removal.

4.3 Feature Extraction

The following algorithms were applied to both the AMIGOS and DEAP datasets. They were chosen based on their effectiveness, simplicity, and computational speed, according to prior studies. The software used for this stage was MATLAB 2015.

The feature extraction process was performed in the EEG signals on specific frequency bands. This is why the feature extraction process was preceded by an additional filtering part. Both the AMIGOS and DEAP signals were subjected to band pass FIR filters to extract the beta (15-30 Hz) and gamma (>30 Hz) frequencies, since, according to most of the works consulted, features taken from these frequencies generate the best accuracies in terms of emotion recognition. However, because there are still many researchers who chose to resort to alpha waves (8-15Hz), these were also extracted.

The Hjorth parameters and spectral entropy were calculated using these filtered signals. The wavelet and IMF based features were calculated using the signals obtained after the pre-processing step. This was done, because the algorithms applied are already band-limited.

Since all of the video may not be representative of the emotion it is supposed to trigger, it is best to consider smaller portions of it. As such we chose to divide the signals into 4s epochs and applied the following feature extraction methods within a 50% overlapping window. This was done based on the results obtained in [14].

4.3.1 Hjorth Parameters

In 1970 Bo Hjorth introduced a time domain signal processing technique which gives us insight into the statistical properties of a signal [77]. The Hjorth parameters, activity, mobility and complexity (Table 4.2), have been extensively used as feature extraction methods for EEG, in epilepsy detection [78] and BCI research works [79].

Hjorth Parameter	Mathematical Formula
Activity	$var(y(t))$
Mobility	$\sqrt{\frac{var(y'(t))}{var(y(t))}}$
Complexity	$\frac{mobility(y'(t))}{mobility(y(t))}$

Table 4.2- The Hjorth parameters. The signal is represented by y(t) and var(y(t)) as its variance.

Activity is represented by the variance of the time function and can indicate the surface of power spectrum in frequency domain. That is, the value of activity is large if higher frequency components are more common, and low otherwise. The mobility parameter represents the mean frequency or the proportion of standard deviation of the power spectrum. This is defined as the square root of variance of the first derivative of the signal divided by the variance of the signal. The complexity parameter indicates how the shape of a signal is similar to a pure sine wave. The value of complexity converges to 1 as the shape of signal becomes more similar to a pure sine wave.

4.3.2 Spectral Entropy

Generally, entropy is a concept related to uncertainty or disorder. It is a measure of the unpredictability of the state, or equivalently, of its average information content. In the field of information theory, this definition is analogous to the definition in thermodynamics. If we consider x_i to be a random variable and $P(x_i)$ its respective probability, then Shannon entropy or the information entropy can be calculated as follows:

$$H(x) = -\sum_{i=1}^{n} P(x_i) \log_2 P(x_i) \quad \text{Eq 4.1}$$

The spectral entropy of a signal is a measure of its spectral power distribution. The concept is based on the above Shannon entropy. The spectral entropy treats the signal's normalized power distribution in the frequency domain as a probability distribution, and calculates the Shannon entropy of it. Therefore, the Shannon entropy in this context is the spectral entropy of the signal if we consider P(x) to be the probability distribution of a power spectrum, like so:

$$P(x) = \frac{S(x)}{\sum_i S(i)} \quad \text{Eq 4.2}$$

Here, S(x) is the power spectrum, which is equal to the absolute value of the discrete Fourier transform of the signal.

4.3.3 Wavelet Energy and Entropy

Wavelet transform is a spectral analysis technique in which any general function can be expressed as an infinite series of wavelets. The basic idea underlying wavelet analysis consists of expressing a signal as a linear combination of a particular set of functions, obtained by shifting and dilating one single function ψ(t) called a mother wavelet [80]. This expression is accomplished by considering all possible integer translations of ψ(t) and dilation is obtained by multiplying t by a scaling factor, which is usually factors of 2 [74]. Eq 4.3 shows how wavelets are generated from the mother wavelet:

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^{j/2}t - k) \quad \text{Eq 4.3}$$

The DWT (Discrete Wavelet Transform) is derived from the continuous wavelet transform with discrete input. DWT analyses the signal at different frequency bands, with different resolutions by decomposing the signal into a coarse approximation and detail information. DWT employs two sets of functions called scaling functions and wavelet functions, which are related to low-pass and high-pass filters, respectively. The decomposition of the signal into the different frequency bands is merely obtained by consecutive high-pass and low-pass filtering of the time domain signal [80]. Briefly, discrete wavelet transform applies a low pass filter to get the low frequency components and a high pass filter to get the high frequency components [74]. The down-sampled outputs of first high-pass and low-pass filters provide the detail coefficients, D1 and the approximation coefficients, A1, respectively. The first approximation, A1 is further decomposed and this process is continued as shown in Figure 4.4.

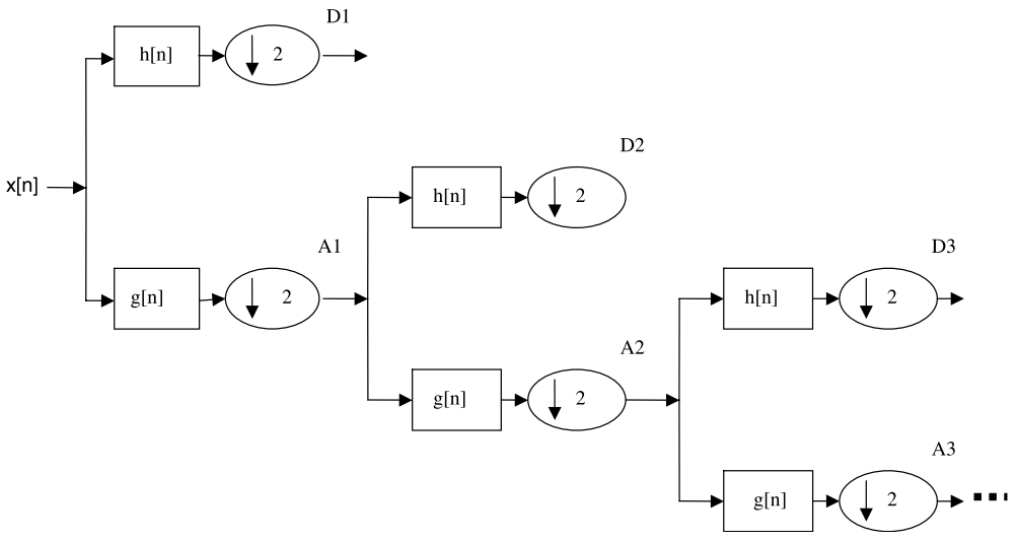

Figure 4.4- Sub-band decomposition of DWT implementation; h[n] is the mother wavelet high-pass filter, g[n] is its low-pass filter mirror [80].

Selection of suitable wavelet and the number of decomposition levels is very important in analysis of signals using the DWT. The number of decomposition levels is chosen based on the dominant frequency components of the signal. The levels are chosen such that those parts of the signal that correlate well with the frequencies necessary for classification of the signal are retained in the wavelet coefficient [80]. For this work, we considered alpha, beta and gamma frequencies, so the number of decomposition levels chosen was 3. D1 would correspond to gamma (32-64Hz), D2 would correspond to beta (16-32Hz) and D3 (8-16Hz) to alpha. The mother wavelet chosen was db4, since it had already proven to generate good results in past works.

Finally, after obtaining the detailed coefficients of the desired bands (decomposition levels) we calculated the energy, by summing the square of the absolute value of these coefficients. The wavelet entropy was calculated in a similar fashion to the spectral entropy.

Both the wavelet energy and wavelet entropy of alpha, beta and gamma (D3, D2, D1) bands were taken as features.

4.3.4 IMF Energy and Entropy

EMD (empirical mode decomposition) is an empirical and data-driven method developed to perform on non-stationary, non-linear, stochastic processes and therefore it is ideally suitable for EEG signal analysis and processing. EMD algorithm decomposes a signal, x[n] into a sum of the band-limited components/functions, c[n] called intrinsic mode functions (IMF) with well-defined instantaneous frequencies. There are two basic conditions to be an IMF: the number of extrema must be equal (or at most may differ by one) to the number of zero crossings; any

point, the mean value of the two envelopes defined by the local maxima and the local minima has to be zero. The general process flow of EMD algorithm is shown below, step by step.

1. Identify all the local extrema.
2. Separately connect all the maxima and minima with natural cubic spline lines to form the upper, u[n], and lower, l[n], envelopes.
3. Find the mean of the envelopes as $z[n] = \frac{[u[n]+l[n]]}{2}$.
4. Take the difference between the data and the mean as the proto-IMF, $h[n] = x[n] — z[n]$.
5. Check the proto-IMF against the definition of IMF and the stoppage criterion to determine if it is an IMF.
6. If the proto-IMF does not satisfy the definition, repeat step1to5on h[n] as many time as needed till it satisfies the definition.
7. If the proto-IMF does satisfy the definition, assign the proto-IMF as an IMF component, c[n].
8. Repeat the operation step 1 to 7 on the residue, q[n]= x[n]—c[n], as the data.
9. The operation ends when the residue contains no more than one extremum.

By the iterative process described above, x(t) can be finally expressed as shown in Eq 4.4.

$$x(t) = \sum_{n=1}^{L} imf_n + r \qquad \text{Eq 4.4}$$

It is a linear combination of IMF components and the residual part r. EMD works as an adaptive high pass filter. It shifts out the fastest changing component first and as the level of IMF increases, the oscillation of IMF becomes smoother. Each component is band-limited, which can reflect the characteristic of instantaneous frequency [56]. The first IMFs contain information in the high frequency spectrum, while the last IMFs contain information within the lowest frequency spectrum.

For this work, we focused on the first IMF, which roughly contains information within the gamma frequency range, the second IMF which contains the beta frequency spectrum and the third IMF which contains alpha band [81]. Then we proceeded to calculate the energy and entropy of the IMFs in the same way as previously described.

Figure 4.5 is a representation of the entire process, from raw EEG to obtaining the features which were later used for regression. Once again, I remind the reader that, although Figure 4.5 specifies only the AMIGOS dataset, the EEG signals from DEAP were processed in the same way. The sole exception, is the pre-processing stage which was skipped, since we chose to use the already processed signals.

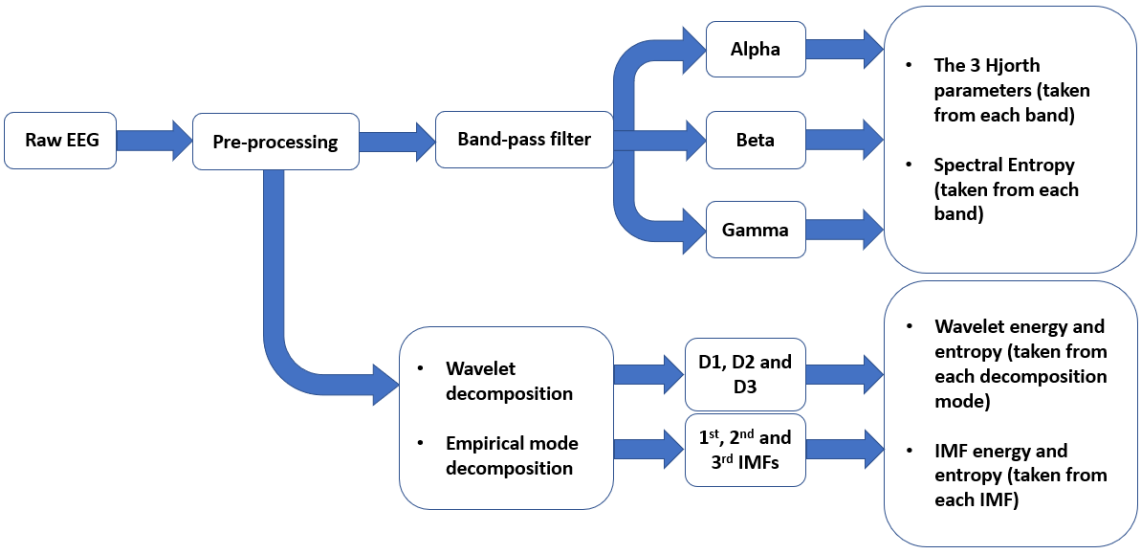

Figure 4.5- This block diagram represents the various stages from pre-processing the raw EEG to obtaining the final features. The pre-processing was the adaptive filtering and wavelet thresholding approach described earlier. Then, the signals were band-pass filtered to obtain the alpha, beta and gamma frequencies and from then the 3 Hjorth parameters were calculated as well as the spectral entropy. A wavelet decomposition was also applied to the pre-processed signals in addition to empirical mode decomposition. The wavelet energy and entropy were calculated from D1, D2 and D3 decomposition modes, and IMF power and entropy were calculated from the first three IMFs.

4.4 Regression

After extracting the features from the EEG signals, we attributed a value of valence/arousal to each of them according to the self-assessment present in the dataset. These values were normalized between -0.5 to 0.5 to compare with works done by other members of the team. Since the features were taken from epochs, the same values of valence/arousal were attributed to every epoch that belonged to the same video. At a later stage of this work, for comparison purposes, we decided to perform the regression, not with data taken from epochs, but rather considering the whole video. In each batch of tests, the objective was to create a model to estimate valence, and another to estimate arousal.

To evaluate the estimators' accuracy, we used three measures: the mean absolute error (MAE), Pearson correlation coefficient (PCC) and root-mean-square error (RMSE). MAE measures the average magnitude of the errors in a set of predictions, without considering their direction. RMSE also measures the average magnitude of the error, but gives a relatively high weight to large errors. In our case, both metrics express the average model prediction error from

0 to 1. PCC measures the linear correlation between the ground-truth and prediction labels. For MAE and RMSE the lower the value the better, while for PCC the closer to 1 the better. The formulas for each of the described measures can be found below. The variable y represents the "ground-truth" samples and $\hat{y}$ represents the estimated values.

$$MAE = \frac{\sum_{i=1}^{N} |\hat{y}_i - y_i|}{N} \quad \text{Eq 4.5}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (\hat{y}_i - y_i)^2}{N}} \quad \text{Eq 4.6}$$

$$PCC = \frac{N \sum_{i=1}^{N} (\hat{y}_i y_i) - \sum_{i=1}^{N} \hat{y}_i \sum_{i=1}^{N} y_i}{\sqrt{N \sum_{i=1}^{N} \hat{y}_i{}^2 - (\sum_{i=1}^{N} \hat{y}_i)^2} \sqrt{N \sum_{i=1}^{N} y_i{}^2 - (\sum_{i=1}^{N} y_i)^2}} \quad \text{Eq 4.7}$$

In order to build the most accurate regression model possible, we had to perform several series of tests. The software used was Weka 3.8 [82] and all the tests were done according to a 10-fold cross validation procedure. In the following paragraphs we are going to describe each of the series of tests as well as the feature vectors used. Although, features within the beta and gamma spectrum seem to generate better results according to most of the works consulted, there are still many who choose to rely on alpha-based ones. For this reason, we decided to perform each batch of tests, considering alpha-based features, beta-based features and gamma-based features separately. The tests were also performed once for each dataset, AMIGOS and DEAP. A generalized feature vector for AMIGOS can be seen below in Figure 4.6. For DEAP the feature vector is in every way the same, except the features were taken from 32 electrode channels instead of 14.

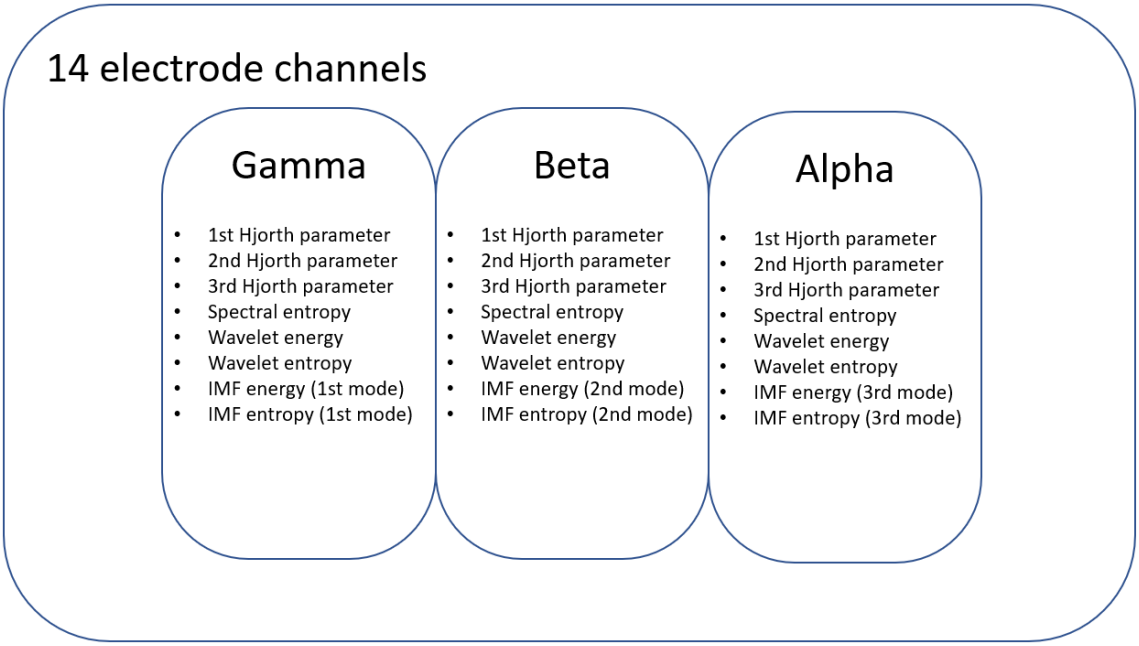

Figure 4.6- The generalized structure of the feature vector used for the tests with AMIGOS dataset. The DEAP feature has the same structure, except the data was taken from 32 electrode channels.

4.4.1 Regression methods

The first batch of tests were meant to determine the best regression methods among the following: AR (additive regression), DT (decision tree), KNN (k-nearest neighbours), LR (linear regression), RF (random forest) and SVR (support vector regression). The last was tested with two different kernels, linear and RBF (radial basis function). All the parameters chosen were the ones already provided by the software. In the end we chose two regressors, RF and KNN, and the rest of the tests were performed only with these two, since they generated the best accuracies (see Table 5.1, in chapter 5). Below you can see a representation of the feature vector used in these tests. It comprises every feature extraction algorithm used, applied to the EEG signals taken from every electrode channel.

The tests were performed three times, one for each frequency band analysed, and this was the procedure taken for the remainder of the work.

$$FV_{AMIGOS} = 14\ channels \times 8\ features = 112\ features$$

$$FV_{DEAP} = 32\ channels \times 8\ features = 256\ features$$

4.4.2 Gender Specificity

The next batch of tests were performed to check for any differences in the predictive accuracy, when separating the participants by gender. Since there are a few studies that report differences in emotional processing between men and women [83] [84], we thought it would be a reasonable hypothesis to test. For these tests we used the following feature vectors:

$$FV_{AMIGOS} = 14\ channels \times 8\ features = 112\ features$$

$$FV_{DEAP} = 32\ channels \times 8\ features = 256\ features$$

4.4.3 Brain Asymmetry

The brain asymmetry concept has been widely used for emotion recognition tasks, particularly for valence prediction. Here we tested two types of asymmetry for valence and arousal estimation: differential and rational. The first was calculated by considering the difference in feature value between electrode channels of opposing hemispheres and only homologue channels were considered (eg. F3-F4, F7-F8, T7-T8…). The rational asymmetry was calculated by dividing the feature values along the same homologue channels. Then we considered both of these asymmetries at the same time.

$$FV_{AMIGOS} = 7\ asymmetry\ channels \times 8\ features \times 1|2\ assymetries = 56|112\ features$$

$$FV_{DEAP} = 14\ asymmetry\ channels \times 8\ features \times 1|2\ assymetries = 112|224\ features$$

4.4.4 Non-asymmetry features plus asymmetry features

Here we tested all the features extracted combined with the asymmetry ones, for the estimation of valence/arousal. Essentially, this model would possess the maximum number of available features.

$$FV_{AMIGOS} = 14\ channels \times 8\ features + (7\ asymmetry\ channels \times 8\ features \times 2\ assymetries) = 224\ features$$

$$FV_{DEAP} = 32\ channels \times 8\ features + (14\ asymmetry\ channels \times 8\ features \times 2\ assymetries) = 480\ features$$

4.4.5 Feature extraction within the whole video

In this scenario, the two best regressors were used to estimate valence/arousal, using the same features as in the other cases, except the asymmetry ones. However, the features used were taken, not in 4s epochs, but rather the whole video. Since the self-assessment annotation gives information related only to the participants' initial and final affective states and not for specific instants during the videos, we thought it might be useful to consider the videos as a whole. This test was only performed using AMIGOS dataset.

$$FV_{AMIGOS} = 14\ channels \times 8\ features = 112\ features$$

4.4.6 Individual electrode comparison

In this group of tests, the goal is to assert any differences in valence/arousal prediction considering each electrode individually. Here we created a model for each individual electrode channel.

$$FV_{AMIGOS} = 1\ channels \times 8\ features = 8\ features$$

$$FV_{DEAP} = 1\ channels \times 8\ features = 8\ features$$

4.4.7 Feature extraction comparison

Here we compared the different feature extraction algorithms used on all electrode channels.

$$FV_{AMIGOS} = 14\ channels \times 1\ features = 14\ features$$

$$FV_{DEAP} = 32\ channels \times 1\ features = 32\ features$$

4.4.8 Feature selection and regressor optimization

Only the best feature extraction methods, obtained in the "feature comparison" analysis, were considered for this step. This search was performed by creating a model using the best feature extraction algorithm, during the feature extraction comparison stage, based on the PCC value. Then, we created another model by adding the second-best feature and compared the results with the previous model. We kept adding features until the results no longer improved.

In the following feature vector representations, the number of feature extraction methods used is different, because it was discovered that the optimal features for regression were different, based on the regressor. For RF, the optimal features were the first Hjorth parameter, wavelet energy and IMF energy. However, KNN sometimes required the use of more features, depending on the dataset used and frequency band.

Then attempt to improve the estimation by changing a few parameters of the regressors. The number of iterations (trees) of RF was increased to 500 and the Manhattan distance was used for the KNN space. Weka also allowed us to choose the best k value automatically through a cross validation-based procedure.

The feature vectors below show the number of features used for each machine learning method in the following way: nº of features in RF | nº of features in KNN.

$$FV_{AMIGOS(Gamma)} = 14\ channels \times 3|3\ features = 42|42\ features$$

$$FV_{AMIGOS(Beta)} = 14\ channels \times 3|7\ features = 42|98\ features$$

$$FV_{AMIGOS(Alpha)} = 14\ channels \times 1|7\ features = 14|98\ \text{features}$$

$$FV_{AMIGOS(Beta+Gamma)} = 14\ channels \times 4|4\ features = 56|56\ features$$

$$FV_{DEAP(Gamma)} = 32\ channels \times 3|3\ features = 96|96\ features$$

$$FV_{DEAP(Beta)} = 32\ channels \times 2|7\ features = 64|224\ features$$

$$FV_{DEAP(Alpha)} = 32\ channels \times 6|7\ features = 192|224\ \text{features}$$

$$FV_{DEAP(Beta+Gamma)} = 32\ channels \times 4|12\ features = 128|384\ features$$

This batch of tests was repeated, but with the addition of the asymmetry-based features (differential only). Once again, it was verified that the optimal feature extraction methods were different according to the regressor.

$$FV_{AMIGOS(Beta+Gamma)} = (14\ channels \times 4|4\ features) + (7\ asymmetry\ channels \times 4|4\ features \times 1\ assymetries) = 84|84\ features$$

$$FV_{DEAP(Beta+Gamma)} = (32\ channels \times 4|12\ features) + (14\ asymmetry\ channels \times 4|12\ features \times 1\ assymetries) = 184|552\ features$$

Additionally, a few *post hoc* pilot tests were performed and showed it is possible to improve the models created using the DEAP dataset using alpha features together with the other frequency bands. The asymmetry tests also showed the importance of the asymmetry concept for valence estimation, and therefore were also used in these tests.

$$FV_{DEAP(RF)\ Arousal} = 32\ channels \times 4\ features = 128\ features$$

$$FV_{DEAP(RF)\ Valence} = 14\ asymmetry\ channels \times 6\ features \times 1\ assymetries = 84\ features$$

$$FV_{DEAP(KNN)\ Arousal} = (32\ channels \times 17\ features) + (14\ asymmetry\ channels \times 17\ features \times 1\ assymetries) = 782\ features$$

$$FV_{DEAP(KNN)\ Valence} = 14\ asymmetry\ channels \times 7\ features \times 1\ assymetries = 98\ features$$

4.4.9 Model validation

The validity of the models was also assessed by testing the best models created earlier, with data from the dataset that was not used for its original creation. That is, we created a model using AMIGOS dataset and tested it using DEAP and vice-versa. We also joined both datasets in an attempt to create a more "generally valid" model. These tests were performed with alpha, beta and gamma band features. Finally, we decided to use all these bands at once with asymmetry and non-asymmetry-based features, since this combination proved to be the best during *post hoc* experiments.

In all these tests we normalized the data and worked with the same number of instances from both datasets in order to prevent any biased results.

$$FV_{train\ with\ AMIGOS\ and\ test\ with\ DEAP(Gamma)} = 14\ channels \times 3|3\ features = 42|42\ features$$

$$FV_{train\ with\ AMIGOS\ and\ test\ with\ DEAP(Beta)} = 14\ channels \times 3|7\ features = 42|98\ features$$

$$FV_{train\ with\ AMIGOS\ and\ test\ with\ DEAP\ (Alpha)} = 14\ channels \times 1|7\ features = 14|98\ features$$

$$FV_{train\ with\ DEAP\ and\ test\ with\ AMIGOS(Gamma)} = 14\ channels \times 3|3\ features = 42|42\ features$$

$$FV_{train\ with\ DEAP\ and\ test\ with\ AMIGOS(Beta)} = 14\ channels \times 2|7\ features = 28|98\ features$$

$$FV_{train\ with\ DEAP\ and\ test\ with\ AMIGOS(Alpha)} = 14\ channels \times 6|7\ features = 84|98\ \ features$$

$$FV_{AMIGOS+DEAP\ (Gamma)} = 14\ channels \times 3|3\ features = 42|42\ features$$

$$FV_{AMIGOS+DEAP\ (Beta)} = 14\ channels \times 3|7\ features = 42|98\ features$$

$$FV_{AMIGOS+DEAP\ (Alpha)} = 14\ channels \times 6|7\ features = 84|98\ features$$

$$FV_{AMIGOS+DEAP\ (Alpha+Beta+Gamma)} = (14\ channels \times 12|17\ features) + (7\ asymmetry\ channels \times 12|17\ features \times 1\ assymetries) = 252|357\ features$$

4.4.10 Classification

After obtaining the optimized model, we were finally ready to classify the valence and arousal values based on the 4 quadrants of the circumplex model of affect. To do that, we simply labelled each pair of values according to H/L (high/low) V/A (valence/arousal) and compared to the estimated values. All the aforementioned procedures were applied to both the AMIGOS and DEAP datasets. Because most works that deal with emotion recognition usually involve classification tasks rather than regression, this step was performed mainly for comparison purposes.

4.5 Summary

The main goal of this work was to compare different elements involved in the emotion recognition task using EEG signals. To accomplish this we used two databases, AMIGOS and DEAP and extracted features from the EEG signals within the alpha, beta and gamma frequency spectrums. These features were the 3 Hjorth parameters, spectral entropy, wavelet energy/entropy and IMF energy/entropy. Once the features were extracted, we built several regression models to estimate valence and arousal levels. These different models were meant to assess, not only how different machine learning models and features can affect the accuracy, but also the brain asymmetry concept, the gender of the individual and the different electrode channels. Since KNN and RF proved to be the superior machine learning approaches by far, they were used throughout the rest of the tests. After achieving the best combination of elements, we built the optimized models and then performed a classification task to compare our work with previous approaches.

In the next chapter, we show the results of each batch of regression tests performed as well as confusion matrixes corresponding to the final classification tasks.

Capítulo 5

Experimental Evaluation

Here we present the results of the various tests mentioned in the previous chapter. This chapter is divided into two sets. The first set represents the results of the V/A estimation. The other set contains the confusion matrixes regarding the classification based on the 4 quadrants of the circumplex model of affect. It is worth mentioning that, although this work serves as a comparative study, certain tests were not performed since it was obvious, they would not improve the results in the attempt to create the best possible model. For example, the search for the best machine learning method was done only once, because the results were quite clear on which were the best machine learning methods to use, on either dataset. The same thing happened with the feature extraction concerning the whole video, that displayed considerably worse accuracy, than the epoch method counterpart.

5.1 Regression results

This section is divided into three subsections. The first, presents the results for the V/A estimation using the AMIGOS dataset. The second subsection contains the results for the V/A estimation using the DEAP dataset. The third subsection shows our attempt to create a "universally" valid model by performing the testing process with a different database than the one that generated the model. This means that in one hand, a model was created using features taken from AMIGOS dataset and tested with DEAP; on the other hand, a model was created using the DEAP dataset and tested with AMIGOS; and finally, both datasets were joined, and a standard 10-fold cross validation procedure was performed. Since one of the objectives was to compare beta, gamma and alpha-based features, the tests were all performed using features from each of these frequency bands separately. Each of the tables shown from now is preceded by a brief description of the test performed and the results obtained. To get a better grasp of the following results, we remind the reader that the labelled V/A values were normalized between -0.5 to 0.5, before training the models.

5.1.1 V/A estimation (AMIGOS)

The next results were obtained using the AMIGOS dataset. By the end of these tests we will have obtained the best combination of features for estimating V/A values, by testing different regressors, gender specificity, symmetric features, electrode channels and the feature extraction algorithms.

The results regarding the use of different regression methods can be seen below in Table 5.1. RF and KNN were the most accurate regressors, and therefore were the only ones used for the remaining tests. The results generated by these approaches are highlighted in bold.

Band	Estimator	Arousal			Valence		
		PCC	MAE	RMSE	PCC	MAE	RMSE
Gamma	AR	0.262	0.180	0.215	0.281	0.232	0.271
	DT	0.477	0.151	0.199	0.422	0.206	0.260
	KNN	**0.490**	**0.131**	**0.224**	**0.472**	**0.173**	**0.289**
	LR	0.188	0.183	0.219	0.199	0.237	0.276
	RF	**0.719**	**0.133**	**0.166**	**0.687**	**0.181**	**0.218**
	SVR (linear)	0.160	0.181	0.222	0.179	0.237	0.279
	SVR (RBF)	0.168	0.181	0.221	0.165	0.238	0.278
Beta	AR	0.254	0.180	0.215	0.270	0.232	0.271
	DT	0.486	0.150	0.198	0.401	0.209	0.264
	KNN	**0.611**	**0.100**	**0.196**	**0.605**	**0.131**	**0.250**
	LR	0.225	0.180	0.217	0.216	0.234	0.275
	RF	**0.725**	**0.131**	**0.165**	**0.683**	**0.181**	**0.218**
	SVR (linear)	0.204	0.180	0.220	0.217	0.234	0.279
	SVR (RBF)	0.213	0.181	0.222	0.221	0.234	0.277
Alpha	AR	0.229	0.181	0.217	0.190	0.239	0.277
	DT	0.353	0.168	0.212	0.255	0.231	0.277
	KNN	**0.519**	**0.125**	**0.218**	**0.478**	**0.170**	**0.288**
	LR	0.173	0.183	0.220	0.161	0.240	0.278
	RF	**0.605**	**0.151**	**0.185**	**0.543**	**0.209**	**0.245**
	SVR (linear)	0.182	0.180	0.221	0.193	0.237	0.278
	SVR (RBF)	0.192	0.180	0.215	0.202	0.238	0.277

Table 5.1- V/A estimation results obtained by comparing various machine learning approaches using AMIGOS dataset. "AR"- Additive Regression; "DT"- Decision Tree; "KNN"- K-Nearest Neighbours; "LR"- Linear Regression; "RF"- Random Forest; "SVR"- Support Vector Regression.

The following table (Table 5.2) contains the results obtained when separating male and female participants. The valence and arousal values of female participants were slightly better estimated than their male counterparts. However, none of the genders were "favoured" when compared to the previous test results, since the highest accuracies were not much different than the ones obtained before considering them separately.

From here on, the best results obtained using KNN and RF for valence and arousal estimation will be highlighted in bold to get a better grasp of the results from the sheer number of tests performed.

Gender	Band	Estimator	Arousal			Valence		
			PCC	MAE	RMSE	PCC	MAE	RMSE
Male	Gamma	KNN	0.400	0.153	0.241	0.367	0.199	0.308
		RF	0.702	0.136	0.168	0.658	0.187	0.222
	Beta	KNN	**0.573**	**0.112**	**0.204**	**0.524**	**0.150**	**0.269**
		RF	**0.707**	**0.133**	**0.166**	**0.661**	**0.186**	**0.221**
	Alpha	KNN	0.508	0.127	0.218	0.451	0.173	0.288
		RF	0.584	0.155	0.187	0.496	0.214	0.248
Female	Gamma	KNN	0.535	0.120	0.217	0.521	0.160	0.286
		RF	**0.739**	**0.124**	**0.161**	0.719	0.174	0.215
	Beta	KNN	**0.620**	**0.096**	**0.196**	**0.621**	**0.129**	**0.256**
		RF	0.727	0.125	0.163	**0.719**	**0.174**	**0.215**
	Alpha	KNN	0.538	0.121	0.217	0.479	0.172	0.299
		RF	0.606	0.146	0.184	0.551	0.214	0.255

Table 5.2- V/A estimation results obtained by considering each gender separately using AMIGOS dataset.

Overall, the asymmetry features did not improve the estimation accuracies. Although neither of the differential or rational asymmetry were superior to one another, using both still generated the best results among this series of tests (Table 5.3).

Asymmetry	Band	Estimator	Arousal			Valence		
			PCC	MAE	RMSE	PCC	MAE	RMSE
Differential	Gamma	KNN	0.180	0.208	0.286	0.223	0.259	0.353
		RF	0.594	0.149	0.185	0.566	0.199	0.238
	Beta	KNN	0.268	0.186	0.270	0.275	0.239	0.340
		RF	0.598	0.147	0.184	0.565	0.199	0.238
	Alpha	KNN	0.305	0.180	0.262	0.284	0.236	0.338
		RF	0.598	0.150	0.184	0.541	0.208	0.244
Rational	Gamma	KNN	0.164	0.212	0.289	0.194	0.267	0.359
		RF	0.591	0.150	0.186	0.581	0.197	0.235
	Beta	KNN	0.280	0.184	0.267	0.289	0.235	0.336
		RF	0.600	0.148	0.184	0.582	0.197	0.236
	Alpha	KNN	**0.436**	**0.145**	**0.235**	0.390	0.199	0.311
		RF	0.597	0.150	0.184	0.538	0.208	0.245
Differential+Rational	Gamma	KNN	0.214	0.262	0.354	0.179	0.209	0.286
		RF	0.600	0.194	0.232	0.624	0.145	0.181
	Beta	KNN	0.282	0.237	0.338	0.277	0.184	0.268
		RF	0.598	0.194	0.233	**0.627**	**0.144**	**0.180**
	Alpha	KNN	0.434	0.147	0.236	**0.402**	**0.197**	**0.308**
		RF	**0.608**	**0.148**	**0.183**	0.549	0.207	0.243

Table 5.3- V/A estimation results obtained when testing the brain asymmetry concept using AMIGOS dataset.

In the next batch of tests (Table 5.4), we employ the feature vector with the maximum size, that is, by using all the available features, both asymmetric and non-asymmetric. It can be seen

that the results did not improve compared to the ones obtained in the first test (see Table 5.1). Alpha-based features did improve the results slightly (see Table 5.1), although still inferior to beta and gamma-based ones.

Band	Estimator	Arousal			Valence		
		PCC	MAE	RMSE	PCC	MAE	RMSE
Gamma	KNN	0.367	0.161	0.250	0.361	0.209	0.318
	RF	0.715	0.133	0.167	**0.682**	**0.182**	**0.218**
Beta	KNN	0.530	0.120	0.216	**0.528**	**0.155**	**0.273**
	RF	**0.718**	**0.131**	**0.165**	0.680	0.182	0.219
Alpha	KNN	**0.545**	**0.119**	**0.212**	0.508	0.162	0.279
	RF	0.616	0.148	0.182	0.554	0.207	0.243

Table 5.4- V/A estimation results obtained when using the brain asymmetry concept, together with non-asymmetry-based features using AMIGOS dataset.

The feature extraction process was originally performed within 4s epochs with 50% overlap. In Table 5.5 we show the results of doing the same procedure only once for the all video. The low accuracies show the clear importance of extracting features within epochs.

Band	Estimator	Arousal			Valence		
		PCC	MAE	RMSE	PCC	MAE	RMSE
Gamma	KNN	0.156	0.218	0.288	0.071	0.308	0.388
	RF	0.329	0.174	0.212	0.366	0.256	0.266
Beta	KNN	**0.254**	**0.204**	**0.272**	**0.140**	**0.288**	**0.370**
	RF	**0.342**	**0.173**	**0.211**	**0.374**	**0.224**	**0.265**
Alpha	KNN	0.197	0.210	0.281	0.133	0.285	0.371
	RF	0.312	0.185	0.224	0.338	0.234	0.289

Table 5.5- V/A estimation results obtained when extracting features from the videos as whole, rather than from epochs.

Next, we compared each electrode channel using all the non-symmetry feature extraction algorithms. Due to the sheer size of the table we decided to move it to the Appendix B section (Table B1). The graphs in Figures 5.1, 5.2, 5.3, 5.4, 5.5 and 5.6 below should help the reader get a better grasp of the results. They show the PCC values for arousal (orange bars) and valence (green bars) estimation. According to the graphs, features taken from the parietal and temporal lobes generated superior accuracies, although by a very slim margin. Therefore, it is impossible to state which electrode channels are the most recommended for V/A estimation.

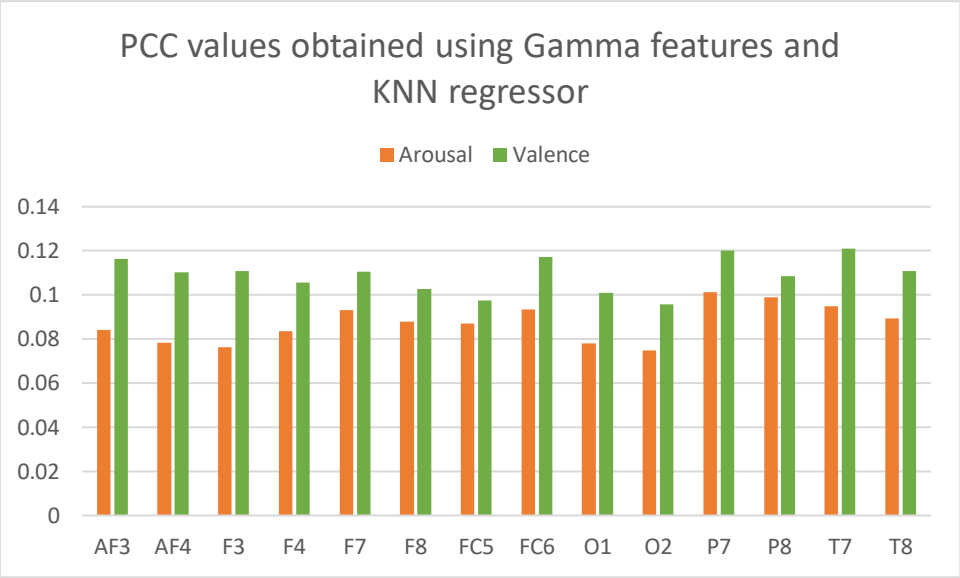

Figure 5.1- PCC values for V/A estimation using KNN regressor and using gamma features for electrode channel comparison.

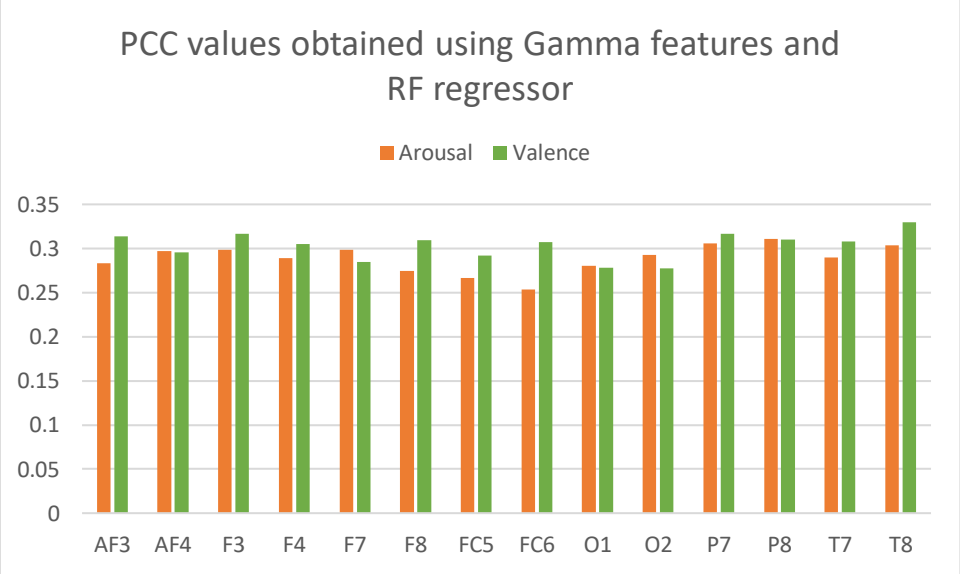

Figure 5.2- PCC values for V/A estimation using RF regressor and using gamma features for electrode channel comparison.

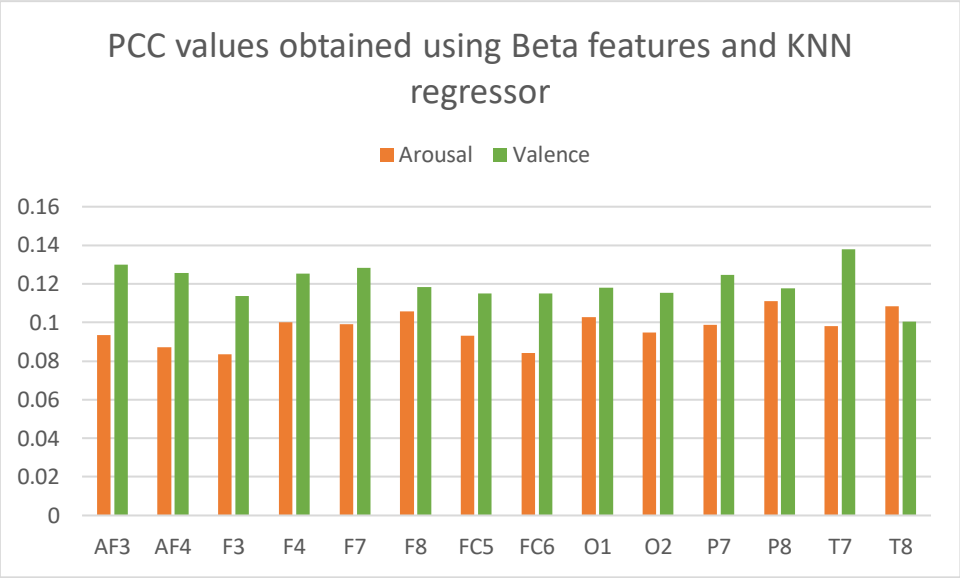

Figure 5.3- PCC values for V/A estimation using KNN regressor and using beta features for electrode channel comparison.

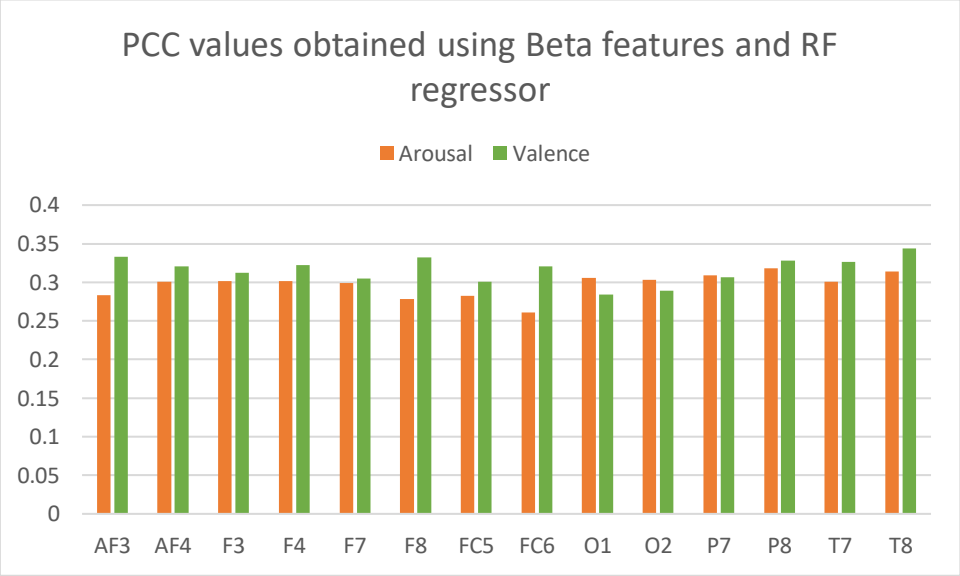

Figure 5.4- PCC values for V/A estimation using RF regressor and using beta features for electrode channel comparison.

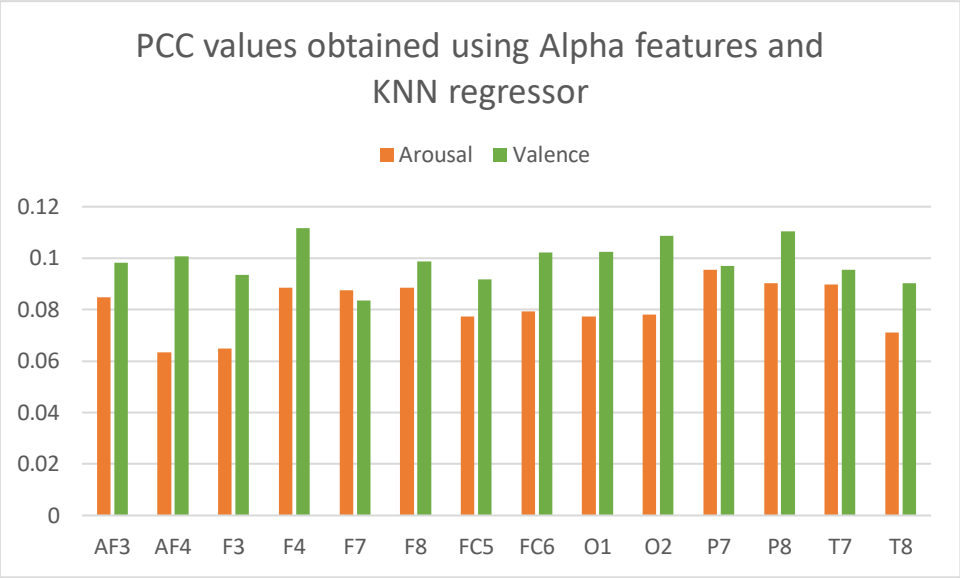

Figure 5.5- PCC values for V/A estimation using KNN regressor and using alpha features for electrode channel comparison.

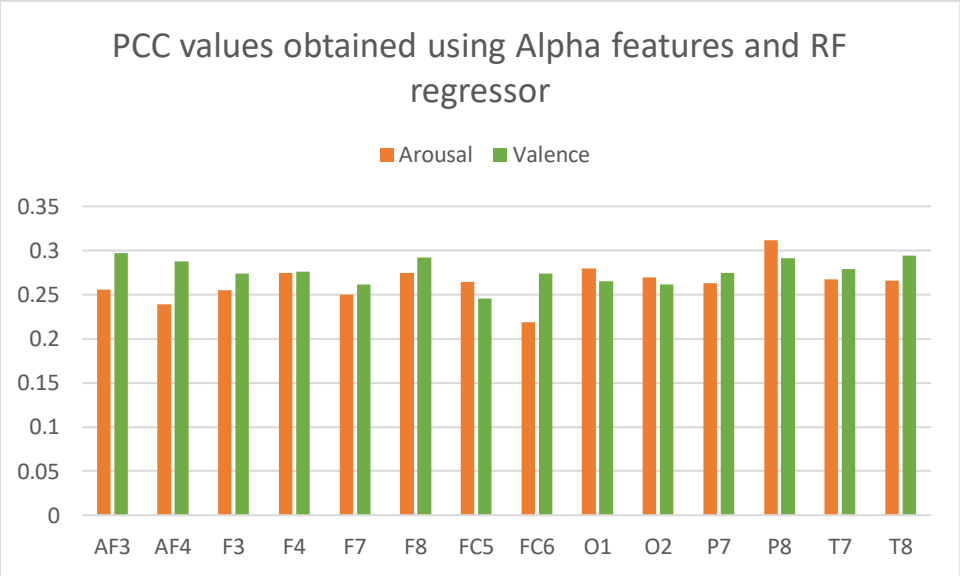

Figure 5.6- PCC values for V/A estimation using RF regressor and using alpha features for electrode channel comparison.

Next, we decided to compare each feature extraction method used. Once again, given the sheer number of instances, we show graphs to help the reader better visualize the data. Figures 5.7, 5.8, 5.9, 5.10, 5.11 and 5.12 show the PCC values for the arousal and valence estimation performed in this modality of tests. The full table can be consulted in Appendix B (table B2). All the results were generated using all the electrode channels, since there were none that were significantly superior in terms of V/A estimation. Once again, beta and gamma-based features were shown to generate the best accuracies.

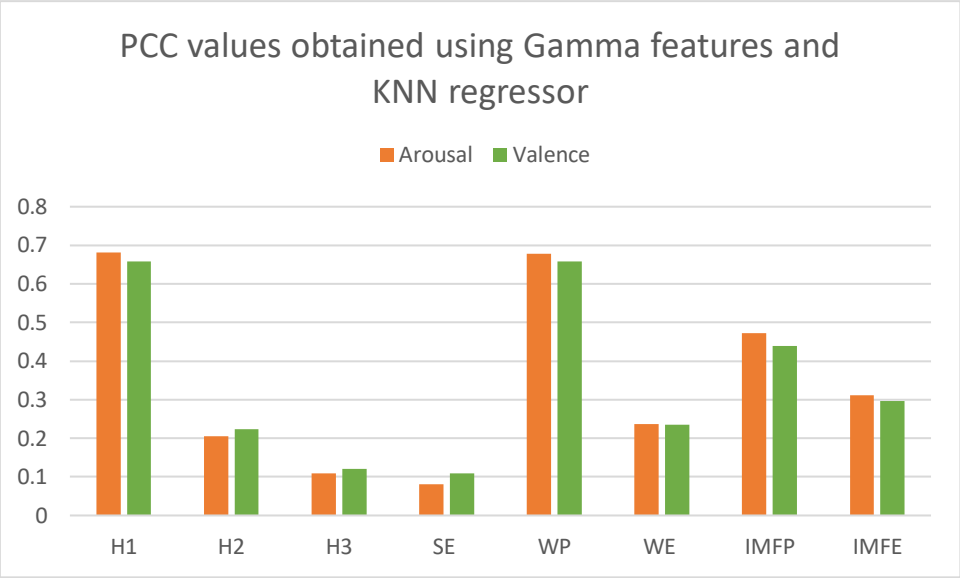

Figure 5.7- PCC values for V/A estimation using KNN regressor and using gamma features for comparing the various feature extraction methods. "H1"- 1st Hjorth Parameter; "H2"- 2nd Hjorth Parameter; 3rd Hjorth Parameter; "SE"- Spectral Entropy; "WP"- Wavelet Energy; "WE"- Wavelet Entropy; "IMFP"- IMF Energy; "IMFE"- IMF Entropy.

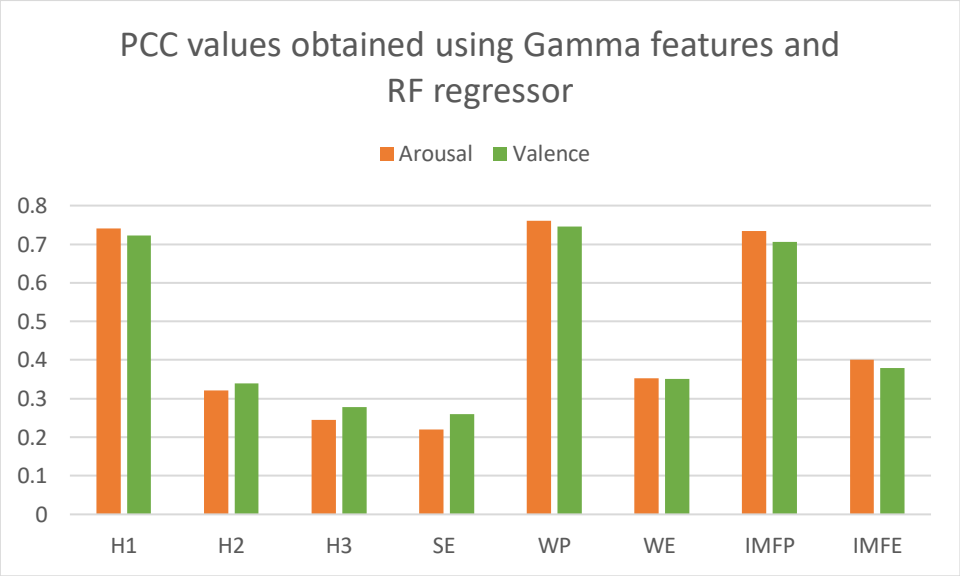

Figure 5.8- PCC values for V/A estimation using RF regressor and using gamma features for comparing the various feature extraction methods. "H1"- 1st Hjorth Parameter; "H2"- 2nd Hjorth Parameter; 3rd Hjorth Parameter; "SE"- Spectral Entropy; "WP"- Wavelet Energy; "WE"- Wavelet Entropy; "IMFP"- IMF Energy; "IMFE"- IMF Entropy.

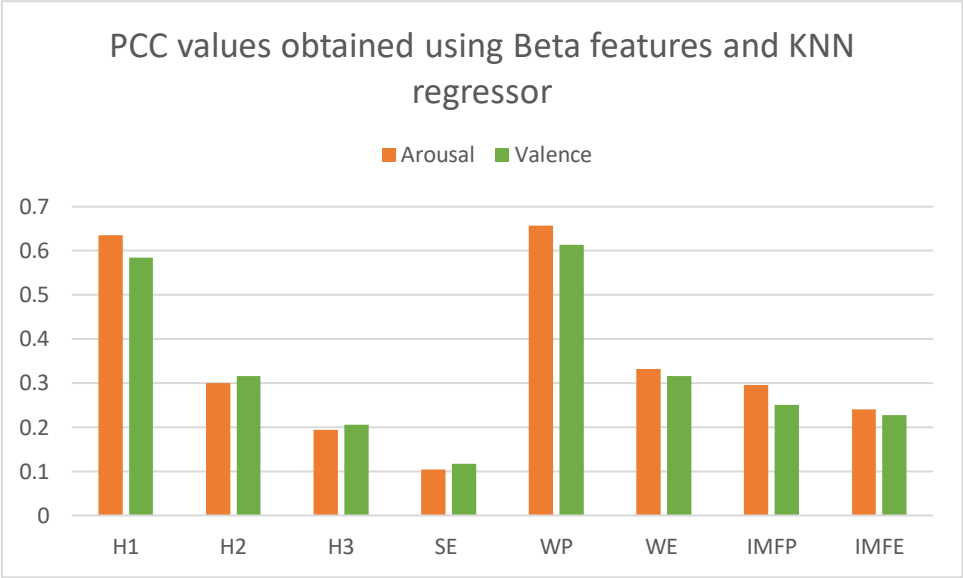

Figure 5.9- PCC values for V/A estimation using KNN regressor and using beta features for comparing the various feature extraction methods. "H1"- 1st Hjorth Parameter; "H2"- 2nd Hjorth Parameter; 3rd Hjorth Parameter; "SE"- Spectral Entropy; "WP"- Wavelet Energy; "WE"- Wavelet Entropy; "IMFP"- IMF Energy; "IMFE"- IMF Entropy.

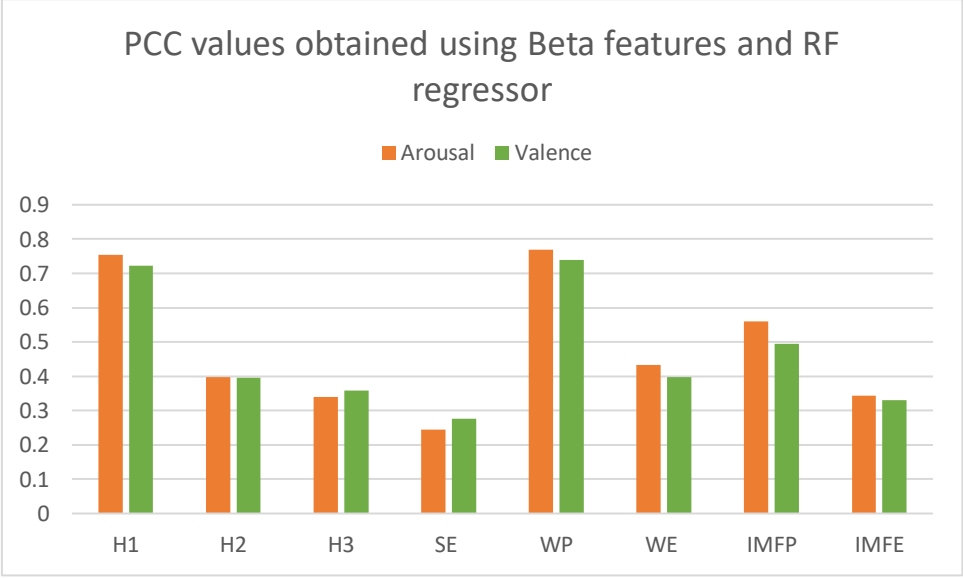

Figure 5.10- PCC values for V/A estimation using RF regressor and using beta features for comparing the various feature extraction methods. "H1"- 1st Hjorth Parameter; "H2"- 2nd Hjorth Parameter; 3rd Hjorth Parameter; "SE"- Spectral Entropy; "WP"- Wavelet Energy; "WE"- Wavelet Entropy; "IMFP"- IMF Energy; "IMFE"- IMF Entropy.

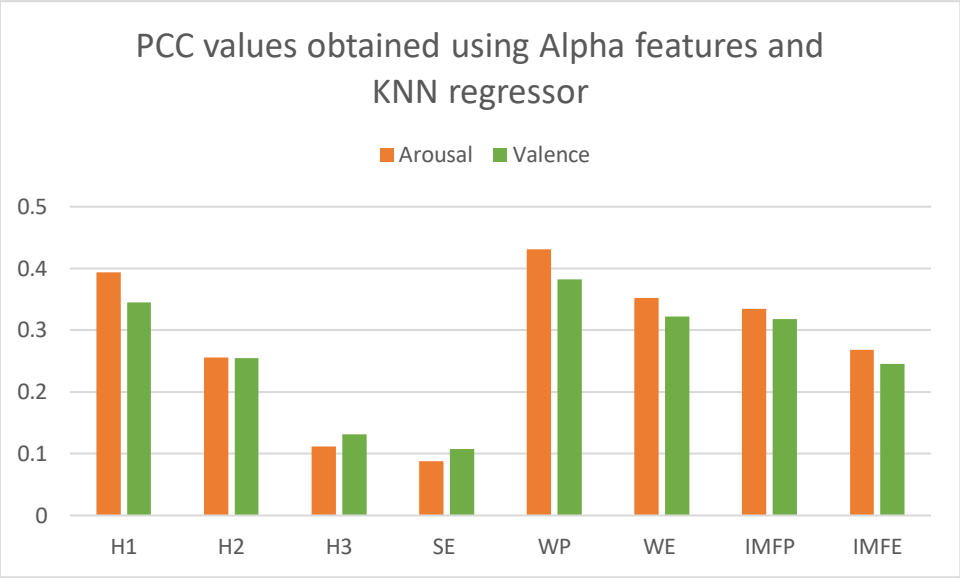

Figure 5.11- PCC values for V/A estimation using KNN regressor and using alpha features for comparing the various feature extraction methods. "H1"- 1st Hjorth Parameter; "H2"- 2nd Hjorth Parameter; 3rd Hjorth Parameter; "SE"- Spectral Entropy; "WP"- Wavelet Energy; "WE"- Wavelet Entropy; "IMFP"- IMF Energy; "IMFE"- IMF Entropy.

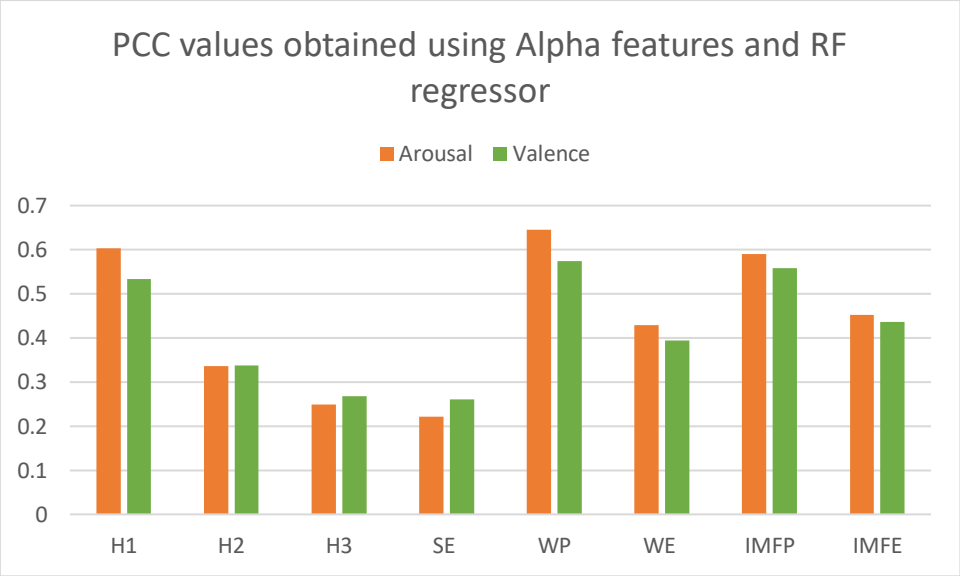

Figure 5.12- PCC values for V/A estimation using RF regressor and using alpha features for comparing the various feature extraction methods. "H1"- 1st Hjorth Parameter; "H2"- 2nd Hjorth Parameter; 3rd Hjorth Parameter; "SE"- Spectral Entropy; "WP"- Wavelet Energy; "WE"- Wavelet Entropy; "IMFP"- IMF Energy; "IMFE"- IMF Entropy.

The final batch of tests consisted of two parts: one in which the best combination of features is used; the other corresponds to the optimization of the regression algorithm. This last part was done by increasing the number of iterations of the RF to 500 (from 100) and by using the Manhattan distance in the neighbour search in KNN (Euclidian distance is standard). The best combination of features used depended on the regressor used and frequency band. For KNN requires more features the lower the frequency band used, while for RF, a combination of Hjorth 1, wavelet energy and IMF energy were used. All the electrode channels were used in the

tests, since removing even one of them, originated worse accuracies. Only the differential asymmetry was considered here, as this one generated the best results in pilot studies. Comparing the results in Table 5.7 with Table 5.8, using the Manhattan distance in KNN was the best course of action.

Band	Estimator	Best feature set	Arousal			Valence		
			PCC	MAE	RMSE	PCC	MAE	RMSE
Gamma	**KNN**	WP, H1, IMFP	0.692	0.082	0.175	0.675	0.110	0.227
	RF	WP, H1, IMFP	0.773	0.115	0.150	0.752	0.158	0.197
Beta	**KNN**	All except SE	0.663	0.091	0.183	0.613	0.129	0.248
	RF	WP, H1, IMFP	0.769	0.113	0.149	0.737	0.160	0.199
Alpha	**KNN**	All except SE	0.519	0.125	0.218	0.478	0.170	0.288
	RF	WP	0.645	0.139	0.175	0.574	0.197	0.236
Gamma+Beta	**KNN**	WP (γ β), H1 (γ β)	**0.753**	**0.068**	**0.157**	**0.710**	**0.098**	**0.214**
	RF	WP (γ β), H1 (γ β)	**0.788**	**0.111**	**0.146**	**0.766**	**0.154**	**0.193**
Gamma+Beta (with asymmetry)	**KNN**	WP (γ β), H1 (γ β)	0.649	0.092	0.187	0.609	0.130	0.248
	RF	WP (γ β), H1 (γ β)	0.782	0.113	0.147	0.758	0.157	0.195

Table 5.6- V/A estimation results obtained when using the best feature set.

Band	Estimator	Arousal			Valence		
		PCC	MAE	RMSE	PCC	MAE	RMSE
Gamma	**KNN**	0.775	0.061	0.150	0.756	0.083	0.196
	RF	0.781	0.115	0.149	0.760	0.157	0.196
Beta	**KNN**	0.742	0.070	0.160	0.704	0.100	0.217
	RF	0.777	0.111	0.148	0.747	0.159	0.199
Alpha	**KNN**	0.604	0.104	0.198	0.570	0.141	0.261
	RF	0.655	0.138	0.174	0.587	0.196	0.234
Gamma+Beta	**KNN**	**0.829**	**0.048**	**0.130**	**0.794**	**0.070**	**0.181**
	RF	**0.794**	**0.111**	**0.145**	**0.770**	**0.155**	**0.193**
Gamma+Beta (with asymmetry)	**KNN**	0.720	0.073	0.166	0.710	0.097	0.214
	RF	0.788	0.112	0.147	0.768	0.156	0.194

Table 5.7- V/A estimation results obtained when using the best feature set and optimizing the estimators.

5.1.2 V/A estimation (DEAP)

The next results were originated using the DEAP dataset. The procedures that were done using AMIGOS dataset were also performed with this dataset with a few exceptions. The initial search for the best regressor was limited to KNN and RF, as it was clear from Table 5.1 that these two were the best choices. The epoch test was also not done since it was clear that the accuracy is much greater when performing the feature extraction in epochs rather than the all video only. It is also worth remembering that the data in DEAP was extracted using a 32-electrode channel set, and each of these will be compared later as well.

In Table 5.9 we can see the results concerning the comparison of accuracies between regressors. Once again, alpha-based features were inferior to the ones extracted within beta and gamma frequencies.

Band	Estimator	Arousal			Valence		
		PCC	MAE	RMSE	PCC	MAE	RMSE
Gamma	KNN	0.409	0.174	0.275	0.387	0.191	0.297
	RF	**0.669**	**0.153**	**0.194**	**0.667**	**0.164**	**0.205**
Beta	KNN	**0.605**	**0.118**	**0.225**	**0.588**	**0.130**	**0.244**
	RF	0.643	0.159	0.199	0.636	0.172	0.213
Alpha	KNN	0.459	0.158	0.263	0.417	0.178	0.290
	RF	0.517	0.177	0.219	0.480	0.197	0.238

Table 5.8- V/A estimation results obtained by comparing KNN and RF machine learning approaches using DEAP dataset.

Table 5.10 below, shows the results obtained when performing the same tests as before, but considering the genders separately. In general, the gender specificity paradigm did not originate better accuracies.

Gender	Band	Estimator	Arousal			Valence		
			PCC	MAE	RMSE	PCC	MAE	RMSE
Male	Gamma	KNN	0.431	0.167	0.266	0.397	0.174	0.277
		RF	**0.641**	**0.163**	**0.200**	**0.607**	**0.167**	**0.206**
	Beta	KNN	**0.639**	**0.109**	**0.215**	**0.605**	**0.115**	**0.225**
		RF	0.601	0.169	0.207	0.570	0.174	0.212
	Alpha	KNN	0.493	0.149	0.254	0.428	0.162	0.271
		RF	0.471	0.187	0.225	0.407	0.195	0.233
Female	Gamma	KNN	0.366	0.187	0.288	0.355	0.210	0.318
		RF	**0.651**	**0.153**	**0.195**	**0.657**	**0.172**	**0.215**
	Beta	KNN	**0.502**	**0.146**	**0.252**	**0.487**	**0.168**	**0.283**
		RF	0.628	0.157	0.199	0.624	0.180	0.223
	Alpha	KNN	0.361	0.184	0.286	0.305	0.217	0.329
		RF	0.474	0.179	0.222	0.412	0.213	0.256

Table 5.9- V/A estimation results obtained by considering each gender separately using DEAP dataset.

The results of using only asymmetry-based features for estimation can be consulted in Table 5.11. The use of differential asymmetry has clearly generated the best accuracies for valence estimation when using the features taken in the alpha spectrum.

Asymmetry	Band	Estimator	Arousal			Valence		
			PCC	MAE	RMSE	PCC	MAE	RMSE
Differential	Gamma	KNN	0.271	0.212	0.305	0.279	0.227	0.322
		RF	0.689	0.148	0.189	0.685	0.158	0.200
	Beta	KNN	0.338	0.195	0.290	0.337	0.209	0.308
		RF	0.674	0.151	0.192	0.672	0.162	0.203
	Alpha	KNN	0.411	0.182	0.275	**0.739**	**0.125**	**0.191**
		RF	0.610	0.166	0.205	**0.862**	**0.120**	**0.153**
Rational	Gamma	KNN	0.447	0.165	0.266	0.448	0.177	0.283
		RF	0.694	0.147	0.188	0.694	0.157	0.194
	Beta	KNN	**0.558**	**0.132**	**0.238**	0.550	0.144	0.255
		RF	0.681	0.151	0.191	0.679	0.162	0.202
	Alpha	KNN	0.410	0.175	0.275	0.366	0.196	0.302
		RF	0.516	0.176	0.218	0.481	0.196	0.237
Differential+Rational	Gamma	KNN	0.422	0.172	0.271	0.423	0.185	0.289
		RF	**0.696**	**0.146**	**0.187**	0.694	0.156	0.198
	Beta	KNN	0.530	0.141	0.245	0.525	0.152	0.262
		RF	0.680	0.150	0.191	0.682	0.161	0.202
	Alpha	KNN	0.472	0.162	0.260	0.672	0.134	0.215
		RF	0.586	0.169	0.209	0.812	0.141	0.176

Table 5.10- V/A estimation results obtained when testing the brain asymmetry concept using DEAP dataset.

The use of asymmetry-based features in conjunction with the non-asymmetry-based ones has produced better results. The increase in accuracy using beta-gamma-based features was slim but was considerable for alpha-based features, as stated in Table 5.12 below.

Band	Estimator	Arousal			Valence		
		PCC	MAE	RMSE	PCC	MAE	RMSE
Gamma	KNN	0.482	0.155	0.258	0.474	0.167	0.276
	RF	**0.689**	**0.149**	**0.189**	**0.687**	**0.159**	**0.200**
Beta	KNN	**0.646**	**0.106**	**0.213**	**0.631**	**0.117**	**0.231**
	RF	0.670	0.154	0.194	0.665	0.166	0.206
Alpha	KNN	0.513	0.147	0.250	0.578	0.145	0.244
	RF	0.564	0.172	0.212	0.740	0.162	0.199

Table 5.11- V/A estimation results obtained when using the brain asymmetry concept, together with non-asymmetry-based features using DEAP dataset.

The next group of tests was meant to analyse each electrode channel individually. As was done previously, we show graphs containing the PCC values for this batch of tests. Figures 5.13, 5.14, 5.15, 5.16, 5.17 and 5.18 can be seen below. Table C1 in Appendix C shows the complete set of results. Once again, it is hard to determine a single or group of electrodes, that was better than the rest in terms of V/A estimation.

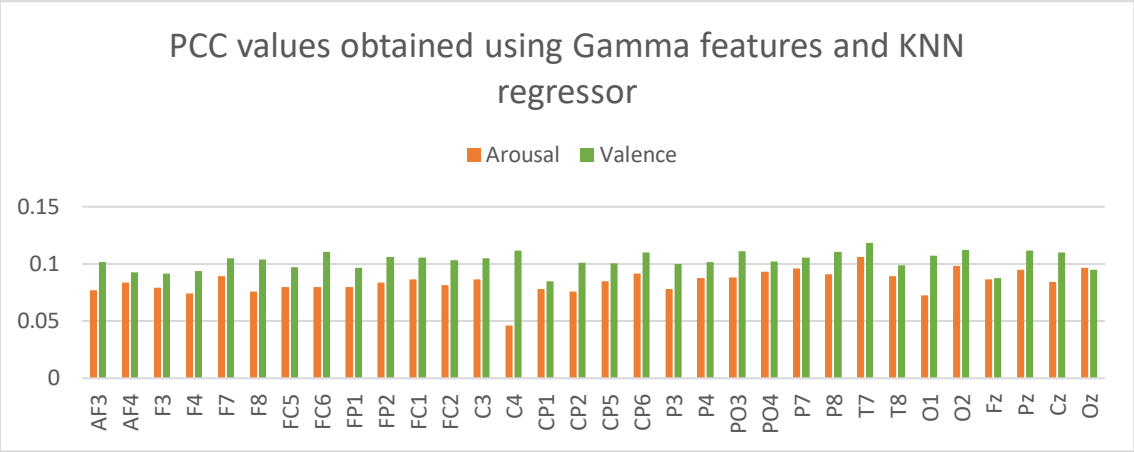

Figure 5.13- PCC values for V/A estimation using KNN regressor and using gamma features for electrode channel comparison.

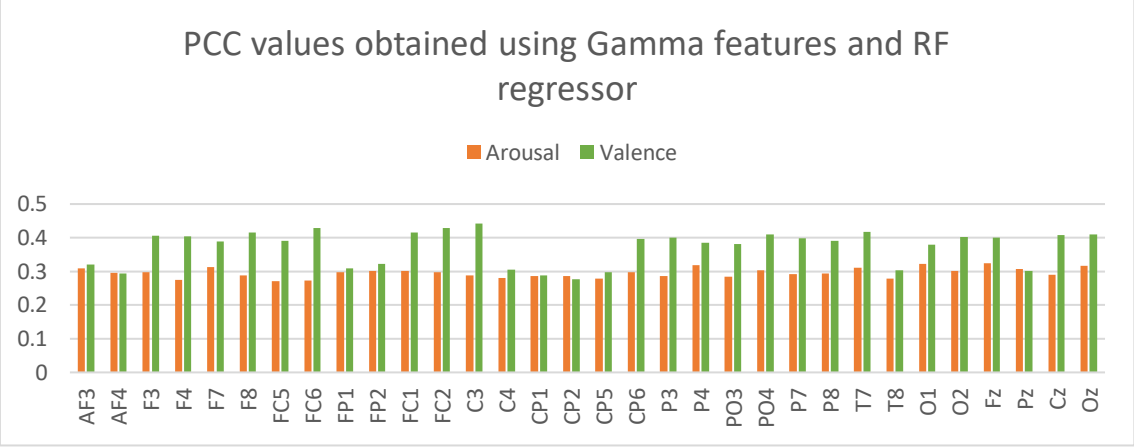

Figure 5.14- PCC values for V/A estimation using RF regressor and using gamma features for electrode channel comparison.

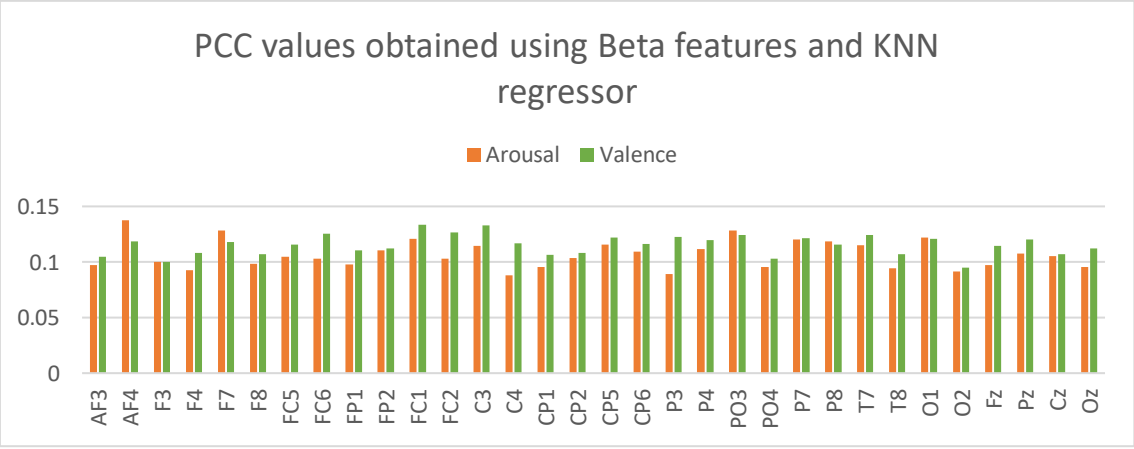

Figure 5.15- PCC values for V/A estimation using KNN regressor and using beta features for electrode channel comparison.

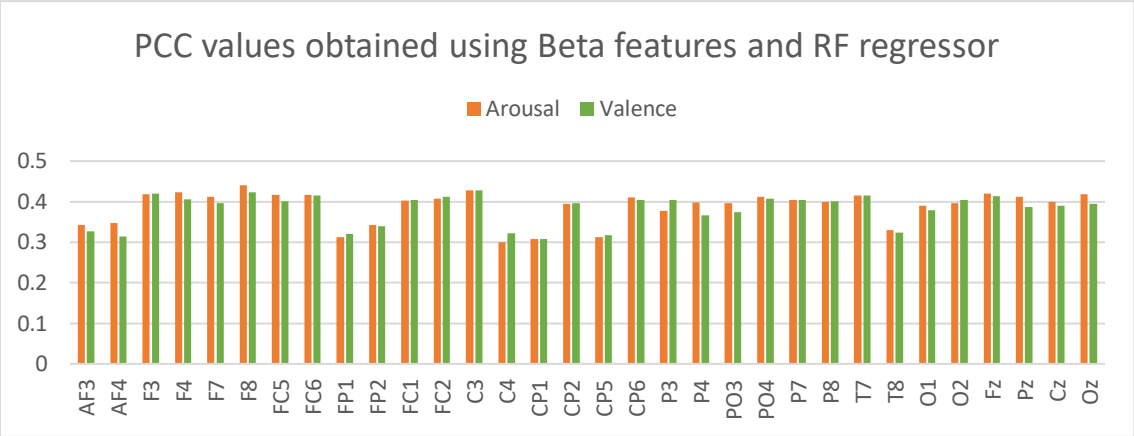

Figure 5.16- PCC values for V/A estimation using RF regressor and using beta features for electrode channel comparison.

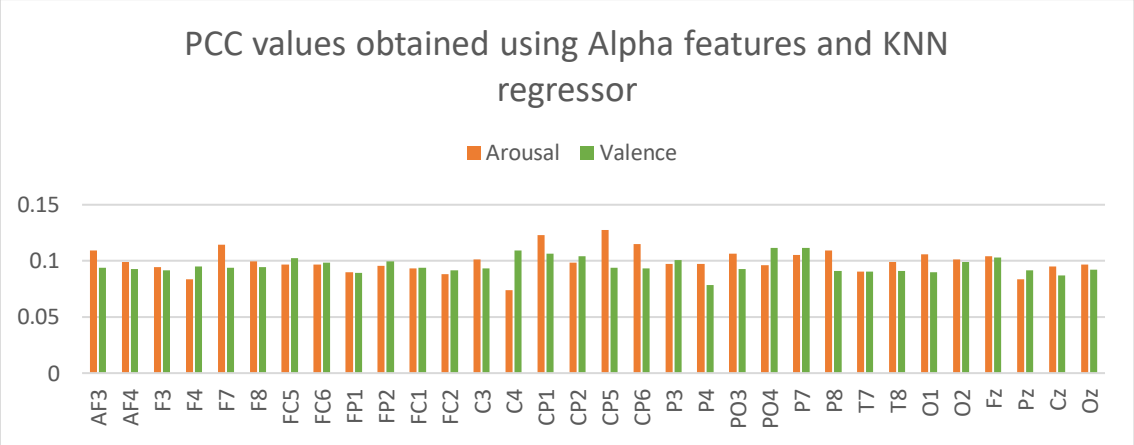

Figure 5.17- PCC values for V/A estimation using KNN regressor and using alpha features for electrode channel comparison.

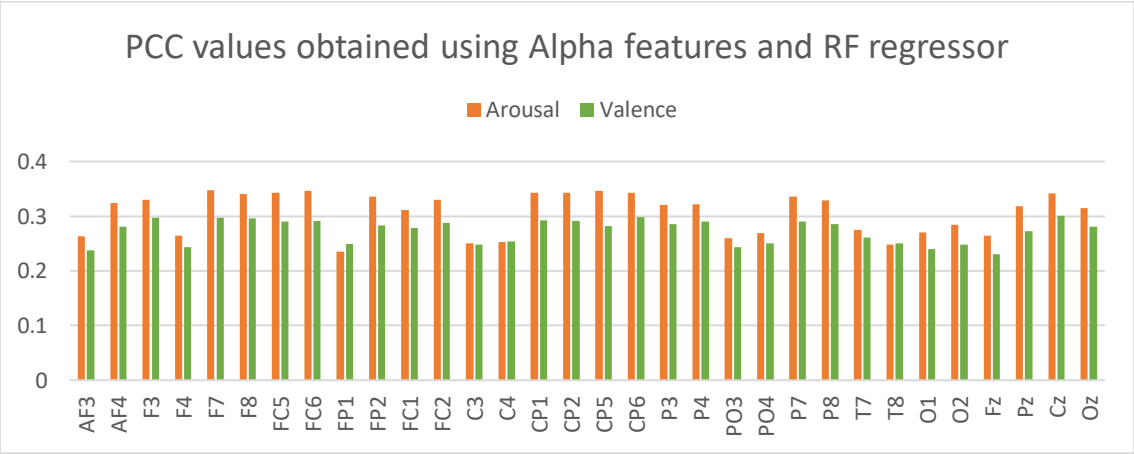

Figure 5.18- PCC values for V/A estimation using RF regressor and using alpha features for electrode channel comparison.

The results of using each feature extraction method algorithm are in Table C2 in Appendix C. Just as done previously, we now show the graphs containing the PCC values for V/A estimation. Figures 5.19, 5.20, 5.21, 5.22, 5.23, and 5.24 can be seen below. Once again, all the electrode channels were used in each test, and beta and gamma-based features originated better accuracies than alpha-based ones.

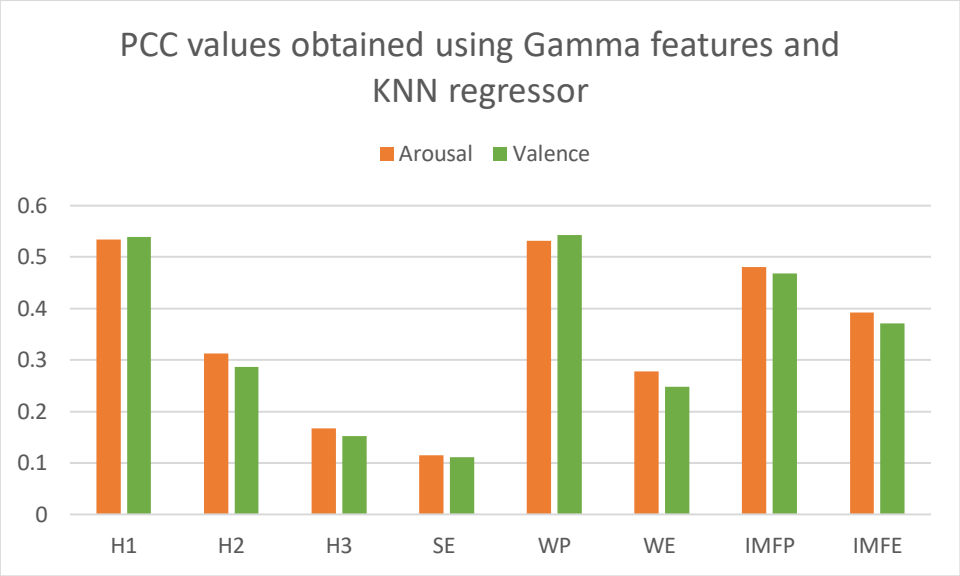

Figure 5.19- PCC values for V/A estimation using KNN regressor and using gamma features for comparing the various feature extraction methods. "H1"- 1st Hjorth Parameter; "H2"- 2nd Hjorth Parameter; 3rd Hjorth Parameter; "SE"- Spectral Entropy; "WP"- Wavelet Energy; "WE"- Wavelet Entropy; "IMFP"- IMF Energy; "IMFE"- IMF Entropy.

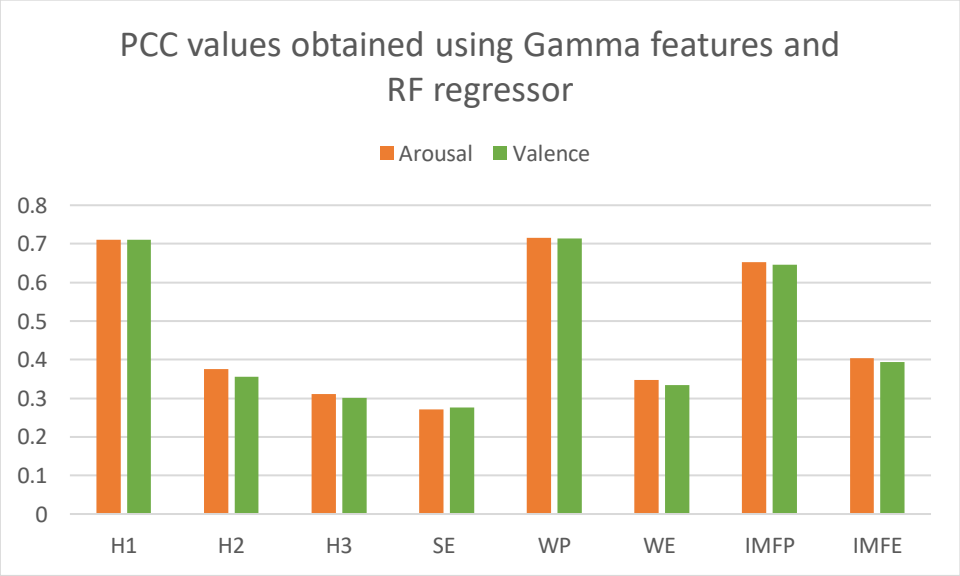

Figure 5.20- PCC values for V/A estimation using RF regressor and using gamma features for comparing the various feature extraction methods. "H1"- 1st Hjorth Parameter; "H2"- 2nd Hjorth Parameter; 3rd Hjorth Parameter; "SE"- Spectral Entropy; "WP"- Wavelet Energy; "WE"- Wavelet Entropy; "IMFP"- IMF Energy; "IMFE"- IMF Entropy.

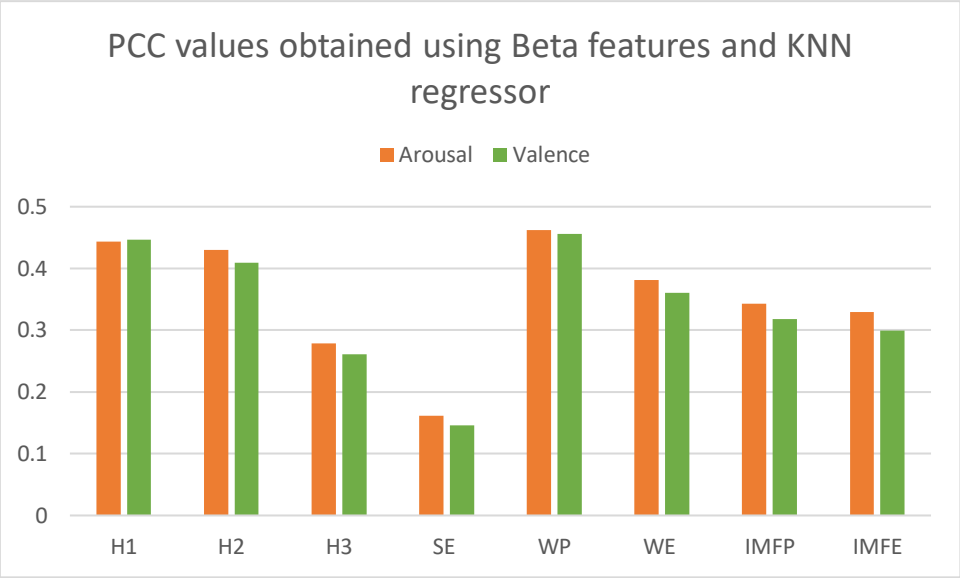

Figure 5.21- PCC values for V/A estimation using KNN regressor and using beta features for comparing the various feature extraction methods. "H1"- 1st Hjorth Parameter; "H2"- 2nd Hjorth Parameter; 3rd Hjorth Parameter; "SE"- Spectral Entropy; "WP"- Wavelet Energy; "WE"- Wavelet Entropy; "IMFP"- IMF Energy; "IMFE"- IMF Entropy.

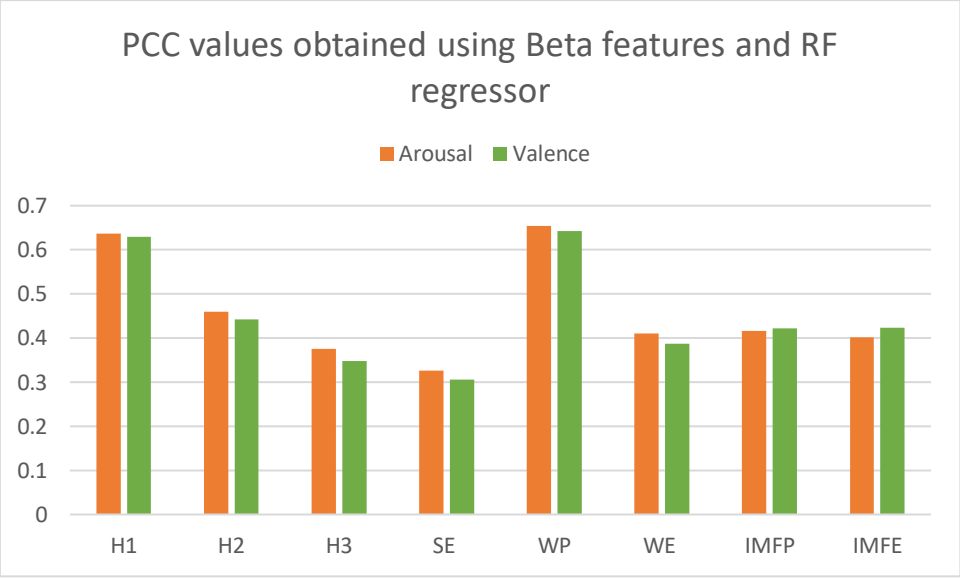

Figure 5.22- PCC values for V/A estimation using RF regressor and using beta features for comparing the various feature extraction methods. "H1"- 1st Hjorth Parameter; "H2"- 2nd Hjorth Parameter; 3rd Hjorth Parameter; "SE"- Spectral Entropy; "WP"- Wavelet Energy; "WE"- Wavelet Entropy; "IMFP"- IMF Energy; "IMFE"- IMF Entropy.

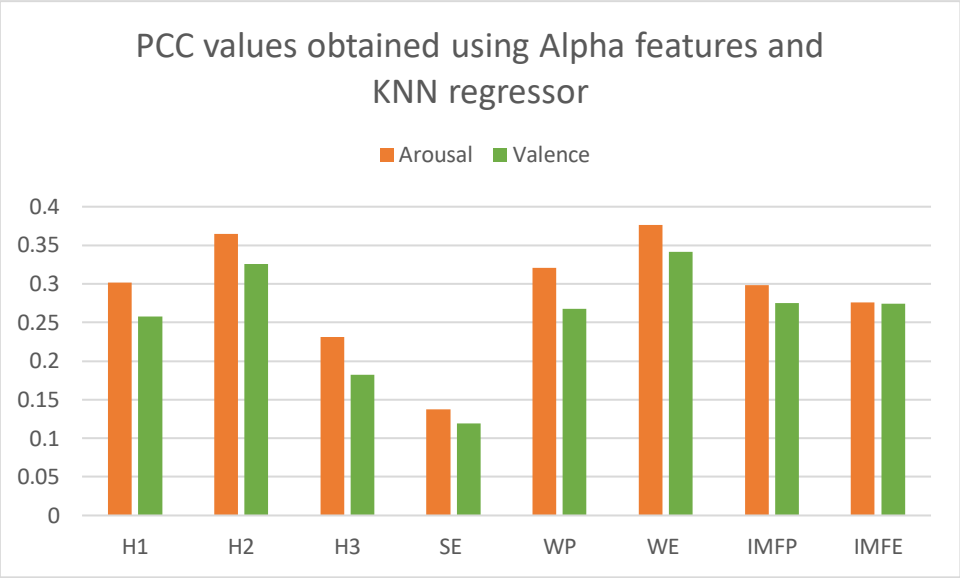

Figure 5.23- PCC values for V/A estimation using KNN regressor and using alpha features for comparing the various feature extraction methods. "H1"- 1st Hjorth Parameter; "H2"- 2nd Hjorth Parameter; 3rd Hjorth Parameter; "SE"- Spectral Entropy; "WP"- Wavelet Energy; "WE"- Wavelet Entropy; "IMFP"- IMF Energy; "IMFE"- IMF Entropy.

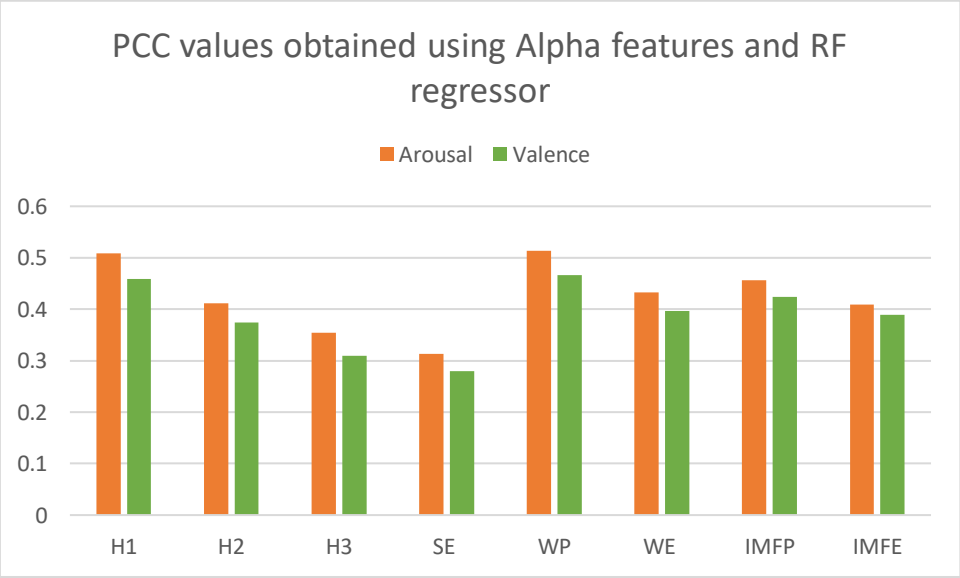

Figure 5.24- PCC values for V/A estimation using RF regressor and using alpha features for comparing the various feature extraction methods. "H1"- 1st Hjorth Parameter; "H2"- 2nd Hjorth Parameter; 3rd Hjorth Parameter; "SE"- Spectral Entropy; "WP"- Wavelet Energy; "WE"- Wavelet Entropy; "IMFP"- IMF Energy; "IMFE"- IMF Entropy.

The best feature extraction algorithms found were the same as in the AMIGOS dataset. So KNN, for the most part, was trained with every feature except spectral entropy, and RF was trained using Hjorth 1, wavelet energy and IMF energy. The optimization parameters were also the same as the ones implemented in the AMIGOS dataset. All the electrode channels were used for the same reason as before. It can be seen in Table 5.13 that the use of asymmetry-based features generated better accuracies for KNN, although the model generated with RF did not

improve. Once again, only differential asymmetry was implemented, since the rational asymmetry was shown to decrease the performance of the model.

Band	Estimator	Best feature set	Arousal			Valence		
			PCC	MAE	RMSE	PCC	MAE	RMSE
Gamma	KNN	WP, H1, IMFP	0.535	0.139	0.244	0.528	0.149	0.260
	RF	WP, H1, IMFP	0.719	0.138	0.180	0.717	0.145	0.190
Beta	KNN	All except SE	0.621	0.114	0.221	0.597	0.126	0.241
	RF	WP, H1	0.669	0.150	0.192	0.661	0.161	0.204
Alpha	KNN	All except SE	0.505	0.147	0.252	0.451	0.167	0.281
	RF	All, except SE and H3	0.543	0.173	0.215	0.505	0.193	0.234
Gamma+Beta	KNN	All except SE	0.672	0.099	0.205	0.652	0.110	0.224
	RF	WP (γ β), H1 (γ β)	**0.722**	**0.138**	**0.180**	**0.719**	**0.147**	**0.190**
Gamma+Beta (with asymmetry)	KNN	All except SE	**0.693**	**0.092**	**0.198**	**0.675**	**0.103**	**0.217**
	RF	WP (γ β), H1 (γ β)	0.717	0.140	0.182	0.713	0.150	0.193

Table 5.12- V/A estimation results obtained when using the best feature set and after optimizing the estimators.

Band	Estimator	Arousal			Valence		
		PCC	MAE	RMSE	PCC	MAE	RMSE
Gamma	KNN	0.604	0.119	0.226	0.596	0.128	0.241
	RF	0.725	0.138	0.179	0.721	0.147	0.190
Beta	KNN	0.667	0.099	0.207	0.648	0.110	0.225
	RF	0.676	0.149	0.191	0.668	0.161	0.204
Alpha	KNN	0.515	0.143	0.249	0.468	0.164	0.279
	RF	0.554	0.172	0.214	0.520	0.192	0.233
Gamma+Beta	KNN	0.718	0.085	0.190	0.700	0.094	0.208
	RF	**0.728**	**0.137**	**0.179**	**0.725**	**0.147**	**0.189**
Gamma+Beta (with asymmetry)	KNN	**0.758**	**0.072**	**0.176**	**0.748**	**0.079**	**0.191**
	RF	0.714	0.142	0.182	0.713	0.152	0.194

Table 5.13- V/A estimation results obtained when using the best feature set and optimizing the estimators.

Unlike with AMIGOS, the models generated with the DEAP dataset appeared to benefit more from the use of alpha-based features. The asymmetry concept was especially useful for estimating valence. For these reasons, we performed an additional batch of tests to see if we could generate better models. For both KNN and RF, valence was estimated using only the differential asymmetry and the best features in the alpha spectrum, according to Table 5.14. For KNN we also used k=5. Arousal estimation with RF was performed in the same way as in Table 5.14, and the alpha features were added to the KNN-based model. The results can be seen below in Table 5.15.

Modality	Estimator	Arousal			Valence		
		PCC	MAE	RMSE	PCC	MAE	RMSE
Non-optimized model	KNN	0.719	0.086	0.186	0.789	0.052	0.130
	RF	0.722	0.138	0.180	0.901	0.097	0.126
Optimized model	KNN	**0.777**	**0.065**	**0.164**	**0.802**	**0.049**	**0.126**
	RF	**0.728**	**0.137**	**0.179**	**0.904**	**0.096**	**0.125**

Table 5.14- V/A estimation results obtained when using the alpha-based features and the brain asymmetry concept.

5.1.3 V/A estimation (DEAP+AMIGOS) and validation

To assess the validity of a model, one must test it using new found data, or in other words, data that was not used in the training process. This was accomplished by creating a model using AMIGOS dataset for the training stage and DEAP for the testing stage. Then we created another model in which DEAP dataset served as training, and AMIGOS as testing. It can be seen below in Table 5.16 that these models are not very accurate, meaning the models created thus far fall short in terms of validity.

One way to fix this is to simply add more data in the training process. As such we created a model using data from both AMIGOS and DEAP datasets. It can be seen in Table 5.16 that the aforementioned approach generated far better accuracies. The models created were optimized using the same parameters as in Table 5.8 and 5.14 along with the same feature extraction algorithms. The models also included the differential asymmetry-based features. Every instance of features was also normalized and we made sure that there were a similar number of them belonging to either dataset to reduce bias.

Modality	Band	Estimator	Arousal			Valence		
			PCC	MAE	RMSE	PCC	MAE	RMSE
Train-AMIGOS Test-DEAP	Gamma	KNN	0.015	0.269	0.346	0.021	0.284	0.362
		RF	0.088	0.198	0.238	0.096	0.201	0.246
	Beta	KNN	0.012	0.243	0.303	0.017	0.289	0.358
		RF	0.086	0.197	0.237	0.129	0.224	0.259
	Alpha	KNN	0.001	0.222	0.273	0.026	0.294	0.353
		RF	0.095	0.207	0.247	0.025	0.235	0.275
Train-DEAP Test-AMIGOS	Gamma	KNN	0.023	0.266	0.344	0.025	0.264	0.342
		RF	0.096	0.185	0.220	0.058	0.272	0.310
	Beta	KNN	0.016	0.259	0.320	0.017	0.317	0.389
		RF	0.090	0.189	0.223	0.037	0.268	0.324
	Alpha	KNN	0.037	0.280	0.340	0.026	0.294	0.353
		RF	0.008	0.194	0.227	0.049	0.245	0.283
Train/Test-AMIGOS+DEAP	Gamma	KNN	0.757	0.069	0.168	0.762	0.075	0.172
		RF	0.787	0.112	0.161	0.787	0.112	0.186
	Beta	KNN	0.735	0.071	0.166	0.749	0.082	0.191
		RF	0.791	0.111	0.147	0.794	0.133	0.172
	Alpha	KNN	0.636	0.098	0.194	0.652	0.114	0.225
		RF	0.660	0.138	0.175	0.664	0.164	0.205
	Alpha+Beta+Gamma	KNN	**0.772**	**0.060**	**0.161**	**0.808**	**0.059**	**0.159**
		RF	**0.794**	**0.111**	**0.145**	**0.794**	**0.133**	**0.171**

Table 5.15- V/A estimation results obtained when assessing how valid were the best models obtained in Table 5.8 and Table 5.14. This table also shows the result of joining both dataset for the training/testing procedure.

5.2 Classification Results

The final step of this work consisted in the actual emotion classification task. Here we assign each V/A value estimated before to the corresponding quadrant in the two-dimensional V/A space. The only models considered were the ones that contained the best combination of features and have also been optimized. In other words, the classification results presented here, were obtained using the models in Tables 5.8 and 5.15. We also show the classification results for the model generated with the combination of AMIGOS and DEAP datasets shown in Table 5.16.

The classification results are all shown in the form of confusion matrixes where each row contains the instances of a predicted class while the columns represent the instances in the real class.

Every matrix shown below contains the classification results using KNN and RF, and it's worth reminding the reader that these models were optimized: KNN was computed using the Manhattan distance and RF was processed with 500 iterations. Valence estimation with KNN in Table 5.28 was performed using k=5.

KNN	HVHA	LVHA	LVLA	HVLA	RF	HVHA	LVHA	LVLA	HVLA
HVHA	**0.823**	0.093	0.031	0.053	**HVHA**	**0.759**	0.127	0.016	0.097
LVHA	0.068	**0.850**	0.031	0.051	**LVHA**	0.149	**0.760**	0.060	0.031
LVLA	0.052	0.069	**0.812**	0.067	**LVLA**	0.054	0.145	**0.630**	0.170
HVLA	0.055	0.058	0.043	**0.845**	**HVLA**	0.153	0.047	0.090	**0.710**
Average accuracy=0.832					**Average accuracy=0.715**				

Table 5.16- Classification using AMIGOS dataset and gamma-based features.

KNN	HVHA	LVHA	LVLA	HVLA	RF	HVHA	LVHA	LVLA	HVLA
HVHA	**0.786**	0.113	0.042	0.059	**HVHA**	**0.747**	0.138	0.021	0.094
LVHA	0.091	**0.815**	0.038	0.056	**LVHA**	0.156	**0.749**	0.062	0.033
LVLA	0.060	0.067	**0.796**	0.077	**LVLA**	0.062	0.133	**0.629**	0.176
HVLA	0.056	0.073	0.053	**0.818**	**HVLA**	0.144	0.061	0.088	**0.707**
Average accuracy=0.804					**Average accuracy=0.708**				

Table 5.17- Classification using AMIGOS dataset and beta-based features.

KNN	HVHA	LVHA	LVLA	HVLA	RF	HVHA	LVHA	LVLA	HVLA
HVHA	**0.708**	0.143	0.059	0.090	**HVHA**	**0.605**	0.214	0.043	0.137
LVHA	0.122	**0.733**	0.056	0.090	**LVHA**	0.233	**0.607**	0.088	0.072
LVLA	0.095	0.109	**0.683**	0.113	**LVLA**	0.097	0.208	**0.456**	0.240
HVLA	0.101	0.108	0.075	**0.716**	**HVLA**	0.193	0.101	0.134	**0.572**
Average accuracy=0.710					**Average accuracy=0.560**				

Table 5.18- Classification using AMIGOS dataset and alpha-based features.

KNN	HVHA	LVHA	LVLA	HVLA	RF	HVHA	LVHA	LVLA	HVLA
HVHA	**0.848**	0.081	0.033	0.038	**HVHA**	**0.768**	0.126	0.014	0.092
LVHA	0.063	**0.880**	0.022	0.035	**LVHA**	0.138	**0.776**	0.058	0.028
LVLA	0.040	0.047	**0.859**	0.053	**LVLA**	0.053	0.138	**0.650**	0.160
HVLA	0.041	0.045	0.038	**0.875**	**HVLA**	0.143	0.048	0.089	**0.720**
Average accuracy=0.866					**Average accuracy=0.738**				

Table 5.19- Classification using AMIGOS dataset and beta+gamma features.

KNN	HVHA	LVHA	LVLA	HVLA	RF	HVHA	LVHA	LVLA	HVLA
HVHA	**0.800**	0.096	0.042	0.061	**HVHA**	**0.771**	0.129	0.015	0.084
LVHA	0.081	**0.814**	0.042	0.063	**LVHA**	0.152	**0.761**	0.058	0.029
LVLA	0.064	0.081	**0.788**	0.068	**LVLA**	0.053	0.143	**0.631**	0.173
HVLA	0.064	0.076	0.047	**0.813**	**HVLA**	0.157	0.050	0.081	**0.711**
Average accuracy=0.805					**Average accuracy=0.729**				

Table 5.20- Classification using AMIGOS dataset and beta+gamma features (asymmetry-based and non-asymmetry based).

KNN	HVHA	LVHA	LVLA	HVLA	RF	HVHA	LVHA	LVLA	HVLA
HVHA	**0.744**	0.097	0.083	0.077	**HVHA**	**0.711**	0.114	0.046	0.128
LVHA	0.147	**0.703**	0.081	0.069	**LVHA**	0.181	**0.663**	0.107	0.050
LVLA	0.142	0.098	**0.678**	0.083	**LVLA**	0.111	0.125	**0.584**	0.179
HVLA	0.147	0.080	0.109	**0.663**	**HVLA**	0.193	0.057	0.135	**0.615**
Average accuracy=0,697					**Average accuracy=0.643**				

Table 5.21- Classification using DEAP dataset and gamma-based features.

KNN	HVHA	LVHA	LVLA	HVLA	RF	HVHA	LVHA	LVLA	HVLA
HVHA	**0.774**	0.085	0.077	0.065	**HVHA**	**0.672**	0.133	0.052	0.143
LVHA	0.131	**0.738**	0.069	0.062	**LVHA**	0.213	**0.615**	0.105	0.067
LVLA	0.121	0.079	**0.722**	0.078	**LVLA**	0.132	0.148	**0.506**	0.214
HVLA	0.115	0.075	0.086	**0.724**	**HVLA**	0.224	0.065	0.127	**0.584**
Average accuracy=0.739					**Average accuracy=0.594**				

Table 5.22- Classification using DEAP dataset and beta-based features.

KNN	HVHA	LVHA	LVLA	HVLA	RF	HVHA	LVHA	LVLA	HVLA
HVHA	**0.663**	0.135	0.108	0.095	**HVHA**	**0.617**	0.146	0.071	0.166
LVHA	0.202	**0.613**	0.093	0.092	**LVHA**	0.345	**0.460**	0.108	0.087
LVLA	0.195	0.104	**0.588**	0.112	**LVLA**	0.219	0.147	**0.368**	0.266
HVLA	0.177	0.113	0.111	**0.598**	**HVLA**	0.312	0.075	0.102	**0.511**
Average accuracy=0.616					**Average accuracy=0.489**				

Table 5.23- Classification using DEAP dataset and alpha-based features.

KNN	HVHA	LVHA	LVLA	HVLA	RF	HVHA	LVHA	LVLA	HVLA
HVHA	**0.807**	0.075	0.062	0.055	**HVHA**	**0.710**	0.117	0.045	0.128
LVHA	0.108	**0.781**	0.059	0.052	**LVHA**	0.181	**0.666**	0.101	0.053
LVLA	0.102	0.071	**0.766**	0.061	**LVLA**	0.107	0.130	**0.580**	0.184
HVLA	0.103	0.061	0.075	**0.760**	**HVLA**	0.195	0.053	0.130	**0.621**
Average accuracy=0.779					**Average accuracy=0.644**				

Table 5.24- Classification using DEAP dataset and beta+gamma features.

KNN	HVHA	LVHA	LVLA	HVLA	RF	HVHA	LVHA	LVLA	HVLA
HVHA	**0.840**	0.061	0.056	0.043	**HVHA**	**0.708**	0.118	0.046	0.127
LVHA	0.092	**0.814**	0.050	0.044	**LVHA**	0.197	**0.642**	0.105	0.056
LVLA	0.083	0.061	**0.806**	0.049	**LVLA**	0.117	0.123	**0.559**	0.202
HVLA	0.083	0.051	0.066	**0.800**	**HVLA**	0.209	0.056	0.126	**0.609**
Average accuracy=0.815					**Average accuracy=0.630**				

Table 5.25- Classification using DEAP dataset and beta+gamma features (asymmetry-based and non-asymmetry based).

KNN	HVHA	LVHA	LVLA	HVLA	RF	HVHA	LVHA	LVLA	HVLA
HVHA	**0.906**	0.014	0.024	0.056	**HVHA**	**0.773**	0.053	0.015	0.159
LVHA	0.133	**0.798**	0.015	0.054	**LVHA**	0.111	**0.735**	0.129	0.025
LVLA	0.127	0.014	**0.792**	0.067	**LVLA**	0.061	0.174	**0.650**	0.114
HVLA	0.139	0.009	0.025	**0.827**	**HVLA**	0.213	0.035	0.108	**0.644**
Average accuracy=0.865					**Average accuracy=0.701**				

Table 5.26- Classification using DEAP dataset and using alpha-based features and the brain asymmetry concept.

KNN	HVHA	LVHA	LVLA	HVLA	RF	HVHA	LVHA	LVLA	HVLA
HVHA	**0.800**	0.053	0.006	0.141	**HVHA**	**0.833**	0.051	0.008	0.109
LVHA	0.125	**0.795**	0.013	0.067	**LVHA**	0.192	**0.712**	0.057	0.039
LVLA	0.115	0.048	**0.794**	0.043	**LVLA**	0.084	0.107	**0.566**	0.242
HVLA	0.093	0.008	0.088	**0.811**	**HVLA**	0.176	0.031	0.055	**0.738**
Average accuracy=0.800					**Average accuracy=0.743**				

Table 5.27- Classification using a combined set made of the AMIGOS and DEAP datasets using alpha, beta and gamma features.

At the end of this section we show a table (Table 5.29) comparing our classification results with some other works done previously. All of them performed classification according to a 4-class distinction paradigm, such as the present work.

Work	Features	Classifier	Average Accuracy	Dataset
Li, Y et al (2017) [61]	Power spectrum density	CLRNN	75.21%	DEAP
Mehmood, R et al (2017) [57]	3 Hjorth parameters	KNN	76.60%	IAPS[2]
R, Alazrai et al (2018) [66]	Quadratic time-frequency distributions	SVM	75.10%	DEAP
Aguiñaga, A et al (2018) [72]	Wavelet Energy	SVM	81.97%	DEAP
Our work	The 3 Hjorth parameters; Wavelet Energy and Entropy; IMF Energy and Entropy	KNN	86.50%	DEAP
Our work	The 1st Hjorth parameter and Wavelet Energy	KNN	86.60%	AMIGOS

Table 5.28- Comparing the results of our emotion classification with some recent approaches. All of these works performed a 4-class classification.

[2] Unlike AMIGOS and DEAP, this dataset does not contain physiological signals. The pictures in IAPS were used as the means for emotion elicitation, and then the researchers in [57] extracted the EEG signals from the participants as they watched those pictures.

5.3 Summary

In this chapter we presented the results concerning V/A estimation using regression methods. KNN and RF were clearly the best regressors and therefore were the ones chosen to carry on the further analysis. We then tested how other factors besides the regressor can affect the accuracy of V/A estimation. These factors were the gender of the participants, the brain asymmetry concept, the different electrodes used, and the feature extraction algorithms.

In short, creating models for each gender did not improve either's accuracy. The brain asymmetry concept was only useful in the models created using DEAP, especially for valence estimation using alpha-based features. When considering models generated by only one electrode, neither stood out, as the differences were not significant. The best feature extraction algorithms were the ones based on the concept of power (Hjorth 1, wavelet energy and IMF energy). However, the optimal combination, was different for each regressor used and frequency band. While RF was more accurate considering only wavelet energy and the first Hjorth parameter, KNN worked best using all the extracted features except spectral entropy. We also found out that despite our encouraging results, the validity of the models is questionable, as testing with different datasets generated low accuracies.

Then we performed a classification task, to get a better sense of our results when compared with other works. In general, our approaches proved to be on par or even superior to other works.

In the next chapter we discuss the results obtained in detail. We point out the most important findings such as the practices that generated the best results. We also mention the practices that were inconclusive and how we could have changed the protocol in order to shed light over these matters.

Capítulo 6

Discussion

The main goal of this work was to identify the best set of practices which ensured the creation of an accurate model for predicting one's emotional state, defined by their valence and arousal values. According to past works, EEG features extracted within the beta and gamma spectrum generate the best results. However, alpha-based features have also proved useful on a few occasions [51] [69], and as such, they were compared with the beta and gamma-based ones throughout the experiment. In the end we built optimized models for V/A estimation, which were then used to classify emotional states based on the quadrants of the circumplex model of affect, since estimation is not the most common approach. By performing a classification task, we can get a better grasp of the significance of our results by comparing with previous works, which mostly focus on classification.

Our search for the best model for V/A estimation started with the comparison of several different machine learning approaches. All these regressors were chosen based on their regular use and overall efficiency. RF and KNN were the regressors used throughout the work as these were the most accurate by far when compared to the rest. Not only did they generate the highest PCC values and lowest errors, but were also relatively fast at training/testing the models, making them the best choice to perform the next analyses. An interesting result to point out was the fact that KNN regressor generated MAE values considerably lower than RF. However, the latter generated lower RMSE values than the KNN regressors. This trend was recurrent during all the work, except during the optimization test, in which there were a few cases where the optimized KNN (Manhattan distance) produced lower MAE and RMSE than RF. This shows that RF generates models that include outlier instances of data, since these greatly affect the RMSE. Nevertheless, when considering the entirety of errors, which is reflected by the MAE value, we can see that KNN is generally superior to RF. In other words, KNN regressors tend to discriminate outliers in favour of a more robust model, which resulted in a more accurate classification task later.

An important difference between the datasets came after the addition of asymmetry-based features to the model. These did not seem to improve the results when using the AMIGOS

dataset, but resulted in a slight increase in estimation accuracy for the DEAP dataset when using beta and gamma-based features. However, the most notable improvement was in the valence prediction when using alpha-based features, as shown in Table 5.11. This does corroborate the valence hypothesis, which proposes that the pattern of hemispheric dominance depends on the emotional valence of the stimulus [8]. According to it, the left hemisphere is dominant for processing positive emotions whereas the right hemisphere is dominant for processing negative emotions [8]. However, the fact that this result was not replicated by the models using the AMIGOS database does raise a few doubts. This might be due to the number of EEG electrodes used, which was much fewer than the DEAP set, thus leading to a decreased number of features extracted. We remind the reader that the AMIGOS dataset was constructed using a 14-electrode headset, while DEAP was built using a 32-electrode headset.

Although differences in emotional processing between male and female participants were not found, this does not mean they are not present. Creating gender-specific models may not have improved V/A estimation for either gender, but a further analysis based on the features and electrode channels could have unravelled these differences. According to research, women tend to react more strongly to negative-valanced stimuli such as anger or disgust [82] [83] and it is possible that the emotion elicitation was not strong enough to trigger such responses. It is also worth noting that, by creating a model specific for each gender we had to use smaller samples, which might not have been enough to generate accurate models.

It was also difficult to assert any significant differences between accuracies obtained when considering each electrode channel individually. Pilot tests performed before the optimization tasks showed that removing any of the channels from the analyses would result in less accurate estimations, which shows that every channel was significant to the task. A similar result was achieved in [35], where the DEAP database was used for emotion classification, and the maximum accuracy was attained when considering every available electrode channel.

When comparing the various feature extraction algorithms used, there was a clear trend in terms of significance. The first Hjorth parameter, the Wavelet Energy and IMF power generated the best accuracies among all the frequency bands tested. These are features heavily correlated with power spectrum density and all have distributions similar to a typical χ^2. However, this does not mean that the rest of the features were non-significant, since they were also necessary to increase the accuracy of the KNN-based models. The sole exception was the Spectral Entropy, which not only was one of the features that generated the lowest accuracies, but was also the least significant since it dropped the PCC values and raised the errors every time it was included during the pilot tests performed before reaching the best feature combination. Beta and gamma-based features also generated the best accuracies, given the higher PCC values and lower errors produced. This is concordant with most of the works consulted that achieved the best results using features extracted within these frequency bands.

The optimization process consisted mostly in tuning the machine learning methods chosen by changing key parameters. For RF, the number of iterations (trees) was changed from the standard 100 to 500, and the accuracies always improved. Higher iteration numbers would certainly yield slightly better results. However, these were not tested given the time and PC memory required. Moreover, it was expected that the increase in accuracy would not be significant. In [58] it can be seen that increasing the number of iterations any more than a certain value no longer resulted in meaningful improvement in terms accuracy.

As for KNN, k=1 was maintained throughout the work, as increasing k generated lower RMSE but higher MAE and lower PCC, meaning the models were not as "globally" accurate. This result also appeared in [71], where KNN was used for emotion classification and k=1 produced the best results. As for the spatial metrics, the Manhattan distance was chosen for optimization as the Euclidean distance can be quite sensitive to outliers. This resulted in improved accuracies for every tested model. The only instance in which a different value of k generated a more accurate model was in the case of valence estimation using alpha asymmetry-based features, which may suggest a high number of outlier instances of data in that frequency spectrum. Pilot *post hoc* tests were performed and confirmed a moderate increase in accuracy when estimating V/A using alpha features.

Although the results obtained are encouraging, one must be aware of the validity of the models created. In tab 5.16, it can be seen that the models generated with one database are not very accurate at predicting V/A values belonging to the other database. This shows that the models created are more valid for the dataset which originated them, making them less accurate when confronted with new data. However, it is difficult to say if this is due to the differences in the devices used for EEG recording, since the combined set made of both databases generated a model with better results.

Finally, in order to compare our work with previous approaches, we turned the obtained V/A estimations into the corresponding quadrants of the circumplex model of affect, since this is by far the most recurrent method according to the works consulted. The higher accuracies obtained were over 86%, and were obtained using KNN based models. This result becomes even more encouraging if we consider that it was obtained via V/A estimation, rather than direct machine learning classification with the four classes. This means that not only can we accurately classify an individual's emotional state, but also assess the level of its expression by obtaining the exact V/A values associated with it.

Capítulo 7

Conclusions and Future Work

In this work we studied the best combination of features and methods, in order to build the best model possible for estimating people's emotions. We extracted various features from EEG signals for emotion assessment, taken from two online-available databases: AMIGOS and DEAP. Using the valence and arousal values provided by the self-assessments as labels, we then created several models for V/A estimation to test how certain factors can affect the accuracy. We tested different machine learning algorithms, considered potential gender differences in emotional expression, verified the reliability of the brain asymmetry hypothesis, tested each electrode channel individually as well as each feature extraction methods. We also performed the tests considering the three EEG frequency bands, alpha, beta and gamma. The accuracy of the estimations was assessed by means of the Pearson Correlation Coefficient, the Mean Squared Error and the Root Mean Squared Error.

In the end we concluded that KNN and RF were the best machine learning approaches, although each worked best with different feature sets. KNN generated the best accuracies when considering all extracted features, except Spectral Entropy, while RF worked best using only power spectrum density-based features such as Wavelet Energy and the first Hjorth parameter. Features extracted within the beta and gamma frequencies were generally the most accurate. Each electrode channel proved equally significant, since removing any would result in a poorer accuracy. The brain asymmetry concept proved useful for valence estimation, generating a considerable increase in performance, although this was only verified in the models created using the DEAP dataset. Finally, the classification performed later proved to be on par with some of the works done, reaching accuracies above 86%, proving our approach to be quite competitive.

For future work, it would be interesting to explore the gender differences in emotional processing with a higher degree of complexity. The analysis would be similar to the already done procedure, but considering each gender separately. That is, we would attempt to compare the significance of the different electrode channels and the different feature extraction algorithms for each gender.

Another factor worthy of further research is the familiarity of the stimuli. In [60] it was performed an emotion classification task using two datasets, one of them being DEAP. The models were created considering two different profiles: one involving participants who were considered familiar with the stimuli and others not. The classification accuracies achieved, for both datasets, were significantly higher for the models generated using data from the participants who were unfamiliar with the music videos.

Lastly, it would also be interesting to explore the usage of connectivity-based features. Most cognitive processes involve the activation of different cortical areas, and affective expression is no exception. Since, there have been works done over this subject with satisfactory results, it could prove useful as an extension of this work.

References

[1] Alarcão, S. M., & Fonseca, M. J., (2017). “Emotions Recognition Using EEG Signals: A Survey”. *IEEE Transactions on Affective Computing*, *3045*(c), 1–20. https://doi.org/10.1109/TAFFC.2017.2714671

[2] Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., & Yger, F. (2018). “A Review of Classification Algorithms for EEG-based Brain-Computer Interfaces: A 10-year Update”. *Journal of Neural Engineering*, 0–20. https://doi.org/10.1088/1741-2552/aab2f2

[3] Jonathan Wolpaw, Elizabeth Wolpaw, Brain-Computer Interfaces: something new under the horizon, “Brain Computer Interface: Principles and Practice” (pp 5), 2012, Oxford University Press

[4] Dalgleish, Tim. "The emotional brain." *Nature Reviews Neuroscience*, vol. 5, no. 7, 2004

[5] Smith, Craig A. & Lazarus, Richard S. (1990). Chapter 23. Emotion and Adaptation. In L.A. Pervin (Ed.). *Handbook of Personality: Theory and Research.* (pp. 609-637). New York: Guilford.

[6] Ekman, P. (1999). "Basic Emotions". In: T. Dalgleish and M. Power (Eds.). *Handbook of Cognition and Emotion*. John Wiley & Sons Ltd, Sussex, UK. https://doi.org/10.1002/0470013494

[7] R.W. Picard, “Affective Computing”. MIT Press, 1997, pp 14-15

[8] Alves, N. T., Fukusima, S. S., & Aznar-Casanova, J. A. (2008). “Models of brain asymmetry in emotional processing”. *Psychology & Neuroscience*, *1*(1), 63–66. https://doi.org/10.3922/j.psns.2008.1.010

[9] Harmon-Jones, E. (2004). “Contributions from research on anger and cognitive dissonance to understanding the motivational functions of asymmetrical frontal brain activity”. *Biological Psychology*, *67*(1–2), 51–76. https://doi.org/10.1016/j.biopsycho.2004.03.003

[10] Onton, J. (2009). “High-frequency broadband modulation of electroencephalographic spectra”. *Frontiers in Human Neuroscience*, *3*(December), 1–18. https://doi.org/10.3389/neuro.09.061.2009

[11] Koelstra, S. (2012). “Deap: A database for emotion analysis; using physiological signals”. *IEEE Transactions on Affective Computing*, *3*(1), 18–31. https://doi.org/10.1109/T-AFFC.2011.15

[12] Ackermann, P., Kohlschein, C., Bitsch, J. Á., Wehrle, K., & Jeschke, S. (2016). "EEG-based automatic emotion recognition: Feature extraction, selection and classification methods". *2016 IEEE 18th International Conference on E-Health Networking, Applications and Services, Healthcom 2016*. https://doi.org/10.1109/HealthCom.2016.7749447

[13] Atkinson, J., & Campos, D. (2016). "Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers". *Expert Systems with Applications*, *47*, 35–41. https://doi.org/10.1016/j.eswa.2015.10.049

[14] Mohammadi, Z., Frounchi, J., & Amiri, M. (2017). "Wavelet-based emotion recognition system using EEG signal". *Neural Computing and Applications*, *28*(8), 1985–1990. https://doi.org/10.1007/s00521-015-2149-8

[15] Chen, J., Hu, B., Moore, P., Zhang, X., & Ma, X. (2015). "Electroencephalogram-based emotion assessment system using ontology and data mining techniques". *Applied Soft Computing Journal*, *30*, 663–674. https://doi.org/10.1016/j.asoc.2015.01.007

[16] M. Soleymani, J. Lichtenauer, T. Pun and M. Pantic, "A Multimodal Database for Affect Recognition and Implicit Tagging," in *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 42-55, Jan.-March 2012. doi: 10.1109/T-AFFC.2011.25

[17] Zheng, W. L., & Lu, B. L. (2015). "Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks". *IEEE Transactions on Autonomous Mental Development*, *7*(3), 162–175. https://doi.org/10.1109/TAMD.2015.2431497

[18] Miranda Correa, J. A., Abadi, M. K., Sebe, N., & Patras, I. (2018). "AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups". *IEEE Transactions on Affective Computing*, (i), 1–14. https://doi.org/10.1109/TAFFC.2018.2884461

[19] Petrantonakis, P. C., & Hadjileontiadis, L. J. (2011). "A novel emotion elicitation index using frontal brain asymmetry for enhanced EEG-based emotion recognition". IEEE T. *IEEE Transactions on Information Technology in Biomedicine*, *15*(5), 737–746. https://doi.org/10.1109/TITB.2011.2157933

[20] Mehmood, R. M., & Lee, H. J. (2016). "A novel feature extraction method based on late positive potential for emotion recognition in human brain signal patterns". *Computers and Electrical Engineering*, *53*, 444–457. https://doi.org/10.1016/j.compeleceng.2016.04.009

[21] Liu, Y., Sourina, O., & Nguyen, M. K. (2011). "Real-time EEG-based emotion recognition and its applications". *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *6670 LNCS*, 256–277. https://doi.org/10.1007/978-3-642-22336-5_13

[22] Lin, Y. P., Wang, C. H., Jung, T. P., Wu, T. L., Jeng, S. K., Duann, J. R., & Chen, J. H. (2010). "EEG-based emotion recognition in music listening". *IEEE Transactions on Biomedical Engineering*, *57*(7), 1798–1806. https://doi.org/10.1109/TBME.2010.2048568

[23] Wang, X. W., Nie, D., & Lu, B. L. (2014). "Emotional state classification from EEG data using machine learning approach". *Neurocomputing*, *129*, 94–106. https://doi.org/10.1016/j.neucom.2013.06.046

[24] Kothe, C. A., Makeig, S., & Onton, J. A. (2013). “Emotion recognition from EEG during self-paced emotional imagery”. *Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013*, 855–858. https://doi.org/10.1109/ACII.2013.160

[25] Petrantonakis, P. C., & Hadjileontiadis, L. J. (2010). “Emotion Recognition from Brain Signals Using Hybrid Adaptive Filtering and Higher Order Crossings Analysis”. *IEEE Transactions on Affective Computing*, *1*(2), 81–97. https://doi.org/10.1109/t-affc.2010.7

[26] Kroupi, Eleni & Vesin, Jean-Marc & Ebrahimi, Touradj. (2015). Subject-Independent Odor Pleasantness Classification Using Brain and Peripheral Signals. *IEEE Transactions on Affective Computing*. 7. 1-1. 10.1109/TAFFC.2015.2496310.

[27] Keuper, K.; Zwitserlood, P.; Rehbein, M.A.; Eden, A.S.; Laeger, I.; Junghofer, M.; Zwanzger, P.; Dobel, C. “Early prefrontal brain responses to the hedonic quality of emotional words—A simultaneous EEG and MEG study”. *PLoS ONE* 2013, 8.

[28] Novosel, A., Lackner, N., Unterrainer, H. *et al.* “Motivational processing of food cues in anorexia nervosa: a pilot study”. *Eating and Weight Disorders - Studies on Anorexia, Bulimia and Obesity* **19,** 169–175 (2014). https://doi.org/10.1007/s40519-014-0114-7

[29] Chanel, G.; Rebetez, C.; Betrancourt, M.; Pun, T. “Emotion assessment from physiological signals for adaptation of game difficulty”. IEEE Trans. Syst. Man Cybern. Part A Syst. Hum. 2011, 41, 1052–1063.

[30] Sourina, O., & Liu, Y. (2011). “A Fractal-Based Algorithm of Emotion Recognition from EEG Using Arousal-Valence Model”, *BIOSIGNALS 2011 - International Conference on Bio-inspired Systems and Signal Processing.* 209–214. https://doi.org/10.5220/0003151802090214

[31] Zheng, Wei-Long & Zhu, Jia-Yi & Lu, Bao-Liang. (2016). Identifying Stable Patterns over Time for Emotion Recognition from EEG. *IEEE Transactions on Affective Computing*. 10.1109/TAFFC.2017.2712143.

[32] Jirayucharoensak, S., Pan-Ngum, S., & Israsena, P. (2014). “EEG-Based Emotion Recognition Using Deep Learning Network with Principal Component Based Covariate Shift Adaptation”, *2014*, *The Scientific World journal*. https://doi.org/10.1155/2014/627892, 10.1155/2014/627892

[33] Soleymani, M., Asghari-Esfeden, S., Fu, Y., & Pantic, M. (2016). “Analysis of EEG Signals and Facial Expressions for Continuous Emotion Detection”. *IEEE Transactions on Affective Computing*, *7*(1), 17–28. https://doi.org/10.1109/TAFFC.2015.2436926

[34] Mert, A., & Akan, A. (2018). “Emotion recognition from EEG signals by using multivariate empirical mode decomposition”. *Pattern Analysis and Applications*, *21*(1), 81–89. https://doi.org/10.1007/s10044-016-0567-6

[35] Zhang, J., Chen, M., Zhao, S., Hu, S., Shi, Z., & Cao, Y. (2016). “ReliefF-based EEG sensor selection methods for emotion recognition”. *Sensors (Switzerland)*, *16*(10), 1–15. https://doi.org/10.3390/s16101558

[36] Zhang, Y., Zhang, S., & Ji, X. (2018). “EEG-based classification of emotions using empirical mode decomposition and autoregressive model”. *Multimedia Tools and Applications*, 1–14. https://doi.org/10.1007/s11042-018-5885-9

[37] Chen, M., Han, J., Guo, L., Wang, J., & Patras, I. (2015). “Identifying valence and arousal levels via connectivity between EEG channels”. *2015 International Conference on Affective Computing and Intelligent Interaction, ACII 2015*, 63–69. https://doi.org/10.1109/ACII.2015.7344552

[38] Winkler, I., Jäger, M., Mihajlović, V., & Tsoneva, T. (2010). “Frontal EEG asymmetry-based classification of emotional valence using common spatial patterns”. *World Academy of Science, Engineering and Technology*, *45*, 373–378.

[39] Lin, Y. P., Wang, C. H., Wu, T. L., Jeng, S. K., & Chen, J. H. (2008). “Support vector machine for EEG signal classification during listening to emotional music”. *Proceedings of the 2008 IEEE 10th Workshop on Multimedia Signal Processing, MMSP 2008*, 127–130. https://doi.org/10.1109/MMSP.2008.4665061

[40] Lin, Y., Wang, C., Wu, T., Jeng, S., & Chen, J. (2009). “EEG-Based Emotion Recognition in Music Listening: A Comparison of Schemes for Multiclass Support Vector Machine” *IEEE*, 489–492. https://doi.org/978-1-4244-2354-5/09

[41] Murugappan, M., Ramachandran, N., & Sazali, Y. (2010). “Classification of human emotion from EEG using discrete wavelet transform”. *Journal of Biomedical Science and Engineering*, *03*(04), 390–396. https://doi.org/10.4236/jbise.2010.34054

[42] Petrantonakis, P. C., & Hadjileontiadis, L. J. (2012). “Adaptive emotional information retrieval from EEG signals in the time-frequency domain”. *IEEE Transactions on Signal Processing*, *60*(5), 2604–2616. https://doi.org/10.1109/TSP.2012.2187647

[43] Ramirez, R., & Vamvakousis, Z. (2012). “Detecting emotion from EEG signals using the Emotive Epoc device”. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7670 LNAI*, 175–184. https://doi.org/10.1007/978-3-642-35139-6_17

[44] M Murugappan, & Subbulakshmi Murugappan. (2013). “Human Emotion Recognition Through Short Time Electroencephalogram (EEG) Signals Using Fast Fourier Transform (FFT)”, *IEEE 9th International Colloquium on Signal Processing and its Applications*, 8–10. https://doi.org/10.1109/CSPA.2013.6530058

[45] Lee, Y. Y., & Hsieh, S. (2014). “Classifying different emotional states by means of eeg based functional connectivity patterns”. *PLoS ONE*, *9*(4). https://doi.org/10.1371/journal.pone.0095415

[46] Zheng, W. L., Dong, B. N., & Lu, B. L. (2014). “Multimodal emotion recognition using EEG and eye tracking data”. *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2014*, 5040–5043. https://doi.org/10.1109/EMBC.2014.6944757

[47] Bajaj, V., & Pachori, R. B. (2014). “Human emotion classification from eeg signals using multiwavelet transform”. *Proceedings - 2014 International Conference on Medical Biometrics, ICMB 2014*, (Md), 125–130. https://doi.org/10.1109/ICMB.2014.29

[48] Soleymani, M., Pantic, M., & Pun, T. (2015). “Multimodal emotion recognition in response to videos”. *2015 International Conference on Affective Computing and Intelligent Interaction, ACII 2015*, *3*(2), 491–497. https://doi.org/10.1109/ACII.2015.7344615

[49] Vijayan, A. E., Sen, D., & Sudheer A.P. (2015). “EEG-based emotion recognition using statistical measures and auto-regressive modelling”. *Proceedings - 2015 IEEE International Conference on Computational Intelligence and Communication Technology, CICT 2015*, 587–591. https://doi.org/10.1109/CICT.2015.24

[50] Zhang, Y., Ji, X., & Zhang, S. (2016). “An approach to EEG-based emotion recognition using combined feature extraction method”. *Neuroscience Letters*, *633*, 152–157. https://doi.org/10.1016/j.neulet.2016.09.037

[51] Huang, X., Kortelainen, J., Zhao, G., Li, X., Moilanen, A., Seppänen, T., & Pietikäinen, M. (2016). “Multi-modal emotion analysis from facial expressions and electroencephalogram”. *Computer Vision and Image Understanding*, *147*, 114–124. https://doi.org/10.1016/j.cviu.2015.09.015

[52] Bhatti, A. M., Majid, M., Anwar, S. M., & Khan, B. (2016). “Human emotion recognition and analysis in response to audio music using brain signals”. *Computers in Human Behavior*, *65*, 267–275. https://doi.org/10.1016/j.chb.2016.08.029

[53] Thammasan, N., Moriyama, K., & Fukui, K. (2016). “Continuous Music-Emotion Recognition Based on Electroencephalogram”, *IEICE Transactions on Information and Systems* D(4), 1234–1241.

[54] Lan, Z., Sourina, O., Wang, L., & Liu, Y. (2016). “Real-time EEG-based emotion monitoring using stable features”. *Visual Computer*, *32*(3), 347–358. https://doi.org/10.1007/s00371-015-1183-y

[55] Shahabi, H., & Moghimi, S. (2016). “Toward automatic detection of brain responses to emotional music through analysis of EEG effective connectivity”. *Computers in Human Behavior*, *58*, 231–239. https://doi.org/10.1016/j.chb.2016.01.005

[56] Yan, B., Zhang, C., Zeng, Y., Zhuang, N., Tong, L., & Zhang, H. (2017). “Emotion Recognition from EEG Signals Using Multidimensional Information in EMD Domain”. *BioMed Research International*, *2017*, 1–9. https://doi.org/10.1155/2017/8317357

[57] Mehmood, R. M., Du, R., & Lee, H. J. (2017). “Optimal feature selection and deep learning ensembles method for emotion recognition from human brain EEG sensors”. *IEEE Access*, *5*, 14797–14806. https://doi.org/10.1109/ACCESS.2017.2724555

[58] Gauba, H., Kumar, P., Roy, P. P., Singh, P., Dogra, D. P., & Raman, B. (2017). “Prediction of advertisement preference by fusing EEG response and sentiment analysis”. *Neural Networks*, *92*, 77–88. https://doi.org/10.1016/j.neunet.2017.01.013

[59] Torres-Valencia, C., Álvarez-López, M., & Orozco-Gutiérrez, Á. (2017). “SVM-based feature selection methods for emotion recognition from multimodal data”. *Journal on Multimodal User Interfaces*, *11*(1), 9–23. https://doi.org/10.1007/s12193-016-0222-y

[60] Thammasan, N., Moriyama, K., Fukui, K. ichi, & Numao, M. (2017). “Familiarity effects in EEG-based emotion recognition”. *Brain Informatics*, *4*(1), 39–50. https://doi.org/10.1007/s40708-016-0051-5

[61] Li, Y., Huang, J., Zhou, H., & Zhong, N. (2017). “Human Emotion Recognition with Electroencephalographic Multidimensional Features by Hybrid Deep Neural Networks”. *Applied Sciences*, *7*(10), 1060. https://doi.org/10.3390/app7101060

[62] Bajoulvand, A., Zargari Marandi, R., Daliri, M. R., & Sabzpoushan, S. H. (2017). "Analysis of folk music preference of people from different ethnic groups using kernel-based methods on EEG signals". *Applied Mathematics and Computation*, *307*, 62–70. https://doi.org/10.1016/j.amc.2017.02.042

[63] Katsigiannis, S., & Ramzan, N. (2018). "DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals from Wireless Low-cost Off-the-Shelf Devices". *IEEE Journal of Biomedical and Health Informatics*, *22*(1), 98–107. https://doi.org/10.1109/JBHI.2017.2688239

[64] Nakisa, B., Naim, M., Tjondronegoro, D., & Chandran, V. (2018). "Evolutionary computation algorithms for feature selection of EEG-based emotion recognition using mobile sensors". *Expert Systems with Applications 93*, 143–155. https://doi.org/10.1016/j.eswa.2017.09.062

[65] Li, X., Song, D., Zhang, P., Zhang, Y., Hou, Y., & Hu, B. (2018). "Exploring EEG Features in Cross-Subject Emotion Recognition", *Frontiers in Neuroscience* 12(March). https://doi.org/10.3389/fnins.2018.00162

[66] Alazrai, R., Homoud, R., Alwanni, H., & Daoud, M. I. (2018). "EEG-based emotion recognition using quadratic time-frequency distribution". *Sensors (Switzerland)*, *18*(8), 1–32. https://doi.org/10.3390/s18082739

[67] Liu, Y. J., Yu, M., Zhao, G., Song, J., Ge, Y., & Shi, Y. (2018). "Real-time movie-induced discrete emotion recognition from EEG signals". *IEEE Transactions on Affective Computing*, *9*(4), 550–562. https://doi.org/10.1109/TAFFC.2017.2660485

[68] Munoz, R., Olivares, R., Taramasco, C., Villarroel, R., Soto, R., Barcelos, T. S., … Alonso-Sánchez, M. F. (2018). "Using Black Hole Algorithm to Improve EEG-Based Emotion Recognition". *Computational Intelligence and Neuroscience*, *2018*, 1–21. https://doi.org/10.1155/2018/3050214

[69] Clerico, A., Tiwari, A., Gupta, R., Jayaraman, S., & Falk, T. H. (2018). "Electroencephalography amplitude modulation analysis for automated affective tagging of music video clips". *Frontiers in Computational Neuroscience*, *11*(January), 1–13. https://doi.org/10.3389/fncom.2017.00115

[70] Song, T., Zheng, W., Lu, C., Zong, Y., Zhang, X., & Cui, Z. (2019). "MPED: A multi-modal physiological emotion database for discrete emotion recognition". *IEEE Access*, *7*, 12177–12191. https://doi.org/10.1109/ACCESS.2019.2891579

[71] Izquierdo-Reyes, J., Ramirez-Mendoza, R. A., Bustamante-Bello, M. R., Pons-Rovira, J. L., & Gonzalez-Vargas, J. E. (2018). "Emotion recognition for semi-autonomous vehicles framework". *International Journal on Interactive Design and Manufacturing*, *12*(4), 1447–1454. https://doi.org/10.1007/s12008-018-0473-9

[72] Aguiñaga, A. R., & Ramirez, M. A. L. (2018). "Emotional states recognition, implementing a low computational complexity strategy". *Health Informatics Journal*, *24*(2), 146–170. https://doi.org/10.1177/1460458216661862

[73] Abadi, M. K., Subramanian, R., Kia, S. M., Avesani, P., Patras, I., & Sebe, N. (2015). "DECAF: MEG-Based Multimodal Database for Decoding Affective Physiological Responses". *IEEE Transactions on Affective Computing*, *6*(3), 209–222. https://doi.org/10.1109/TAFFC.2015.2392932

[74] Islam, M. K., Rastegarnia, A., & Yang, Z. (2016). “Methods for artifact detection and removal from scalp EEG: A review”. *Neurophysiologie Clinique/Clinical Neurophysiology*, *46*(4–5), 287–305. https://doi.org/10.1016/j.neucli.2016.07.002

[75] Garces Correa, M. A., & Laciar, E. (2012). “Noise Removal from EEG Signals in Polisomnographic Records Applying Adaptive Filters in Cascade”. *Adaptive Filtering Applications*. https://doi.org/10.5772/17219

[76] Gorji, H., Koohpayezadeh, A., & Haddadnia, J. (2013). “Ocular Artifact Detection and Removing from EEG by wavelet families: A Comparative Study”. *Journal of Information Enginnering and Applications*, *3*(13), 39–48.

[77] H. Bo, "EEG analysis based on time domain properties," *Electroencephalography and Clinical Neurophysiology*, vol. 29, no. 3, pp. 306-310, 1970.

[78] Päivinen, N., Lammi, S., Pitkänen, A., Nissinen, J., Penttonen, M., & Grönfors, T. (2005). “Epileptic seizure detection: A nonlinear viewpoint”. *Computer Methods and Programs in Biomedicine*, *79*(2), 151–159. https://doi.org/10.1016/j.cmpb.2005.04.006

[79] Oh, S.-H., Lee, Y.-R., & Kim, H.-N. (2014). “A Novel EEG Feature Extraction Method Using Hjorth Parameter”. *International Journal of Electronics and Electrical Engineering*, *2*(2), 106–110. https://doi.org/10.12720/ijeee.2.2.106-110

[80] Subasi, A. (2007). “EEG signal classification using wavelet feature extraction and a mixture of expert model”. *Expert Systems with Applications*, *32*(4), 1084–1093. https://doi.org/10.1016/j.eswa.2006.02.005

[81] Tsai, F. F., Fan, S. Z., Lin, Y. S., Huang, N. E., & Yeh, J. R. (2016). “Investigating power density and the degree of nonlinearity in intrinsic components of anesthesia EEG by the Hilbert-Huang Transform: An example using ketamine and alfentanil”. *PLoS ONE*, *11*(12), 1–16. https://doi.org/10.1371/journal.pone.0168108

[82] Wekamachine learning tool. Available: http://www.cs.waikato. ac.nz/ml/weka/downloading.html

[83] Lithari, C., Frantzidis, C. A., Papadelis, C., Vivas, A. B., Klados, M. A., Kourtidou-Papadeli, C., … Bamidis, P. D. (2010). “Are females more responsive to emotional stimuli? A neurophysiological study across arousal and valence dimensions”. *Brain Topography*, *23*(1), 27–40. https://doi.org/10.1007/s10548-009-0130-5

[84] Bilalpur, M., Kia, S. M., Chawla, M., Chua, T.-S., & Subramanian, R. (2017). “Gender and Emotion Recognition with Implicit User Signals”, *ICMI '17: Proceedings of the 19th ACM International Conference on Multimodal Interaction* 379–387. https://doi.org/10.1145/3136755.3136790

Appendix A

Ref	Emotion Elicitation	Brain Waves	Channels	Feature extraction	Feature Selection
[39]	Excerpts of emotional music	δ, θ, α, β, and γ	32 electrodes (10-20 system)	ASM12 (power difference at 12 symmetric electrode pairs) and PSD24	Not performed
[40]	Excerpts of emotional music	δ, θ, α, β, and γ	32 electrodes (10-20 system)	ASM12	Not performed
[22]	Music from Oscar winning films. Out of 16 music excerpts, 4 were chosen to elicit the emotions	δ, θ, α, β, and γ	30 electrodes (monopoles) and 12 pairs of electrodes.	Power spectral density (PSD), and power asymmetry (differential and rational)	F-score
[25]	Pictures of Facial Affect (POFA). 60 images, 10 for each emotion	α and β	Fp1 e Fp2 (monopoles); F3 e F4 (dipoles)	EMD (Empirical Mode Decomposition). Genetic algorithm (GA) was applied to IMFs using two different fitness functions. The final feature vector was generated using High Order Crossing (HOC)	Not performed. It's a comparative study
[41]	5 video clips for three emotions (happy, disgust, surprise) and 4 video clips have been selected for two emotions	α, β, and γ	3 different sets: 62 electrodes, 24 electrodes and 8 electrodes	Wavelet transform using 3 decomposition levels, one for each frequency band (D4, D3, D2, for alpha, beta and gamma respectively). Recoursing Energy Efficiency (REE), Logarithmic REE (LREE) and Absolute Logarithmic REE (ALREE) were calculated. Power, standard deviation (SD) and variance were also calculated	Not performed
[38]	48 positive, 48 negative and 16 neutral pictures from a Philips-intern picture database	α	32 electrodes (10-20 system)	PSD in alpha and PSD asymmetry in alpha. CSP (Common Spatial Patterns) was also performed	Not performed
[21]	**2 Sets-** 5 music pieces, 1 for each emotion; 27 music pieces from IADS	Raw EEG (2-42Hz)	FC6 (for arousal estimation) and AF3-F4 asymmetry (for valence estimation)	Fractal Dimension (FD) using Higuchi's method	Not performed
[19]	40 images from IAPS. 10 for each emotion	α and β	Fp1 e Fp2 (monopoles); F3 e F4 (dipoles)	HOC and CC (cross correlation)	Not performed

Ref	Emotion Elicitation	Brain Waves	Channels	Feature extraction	Feature Selection
[30]	**2 Sets-** 5 music pieces, 1 for each emotion; 27 music pieces from IADS	Raw EEG (2-42 Hz)	14 electrodes (10-20 system)	FD (Higuchi and Box counting) from each electrode	Not performed
[11]	DEAP	θ, α, β and γ	32 electrodes (10-20 system)	Power spectrum and Power spectral asymmetry	Fisher Linear Discriminant
[16]	MAHNOB-HCI.	θ, α, β, and γ	32 electrodes (10-20 system)	In addition to power spectral features, the difference between the spectral power of all the symmetrical pairs of electrodes on the right and left hemispheres was extracted. The asymmetry features were extracted from all mentioned bands	Not performed
[42]	IAPS	α and β	Fp1, Fp2 (monopoles), F3, and F4 (dipole)	AsI calculated using: Multidimensional Directed Information (MDI), EMD. In the end, the feature vector was generated using HOC and CC	Not performed
[43]	12 sound stimuli selected from IADS	α and β	AF3, AF4, F3 and F4	valence = power (aF4/bF4 – aF3/bF3)	Not performed
[44]	5 video clips for three emotions (happy, disgust, surprise) and 4 video clips have been selected for two emotions	α, β, and γ. They also used the α to γ spectrum as a whole	62 electrodes (10-10 system)	Spectral Entropy (SE) e Spectral Centroid (SC)	Not performed
[24]	The participants were required to imagine different emotional scenarios	δ, θ, α, β, and γ	124 electrodes	Common Spatial Pattern (CSP)	GLM
[32]	DEAP	θ, lower α, upper α, β, and γ	32 electrodes (10-20 system)	Power spectral density, and power asymmetry with baseline subtracted	PCA (used in two setups)
[23]	12 film clips. Half for eliciting negative emotions and the other half to elicit positive emotions	δ, θ, α, β, and γ	62 electrodes (10-10 system) and 27 asymmetry pairs	PSD followed by differential asymmetry and wavelet entropy. Approximate entropy, Hurst exponent and fractal dimension were also tested	PCA, linear discriminant analysis (LDA), and correlation-based feature selector (CFS).

Ref	Emotion Elicitation	Brain Waves	Channels	Feature extraction	Feature Selection
[45]	Six emotion-eliciting film clips with visual and auditory components retrieved from the Standard Chinese Emotional Film Clips Database to induce positive (an amusing & a surprising film clip), neutral (two "neutral" film clips), or negative (a fear & a disgust film clip) emotions.	θ, α, β, and γ	64 electrodes (10-10 system)	Correlation, coherence, phase synchronization index (PSI)	ANOVA
[46]	15 emotional film clips were selected to elicit three emotions	δ, θ, α, β, and γ	62 electrodes (10-10 system)	(PSD), differential entropy (DE), differential asymmetry (DASM) and rational asymmetry (RASM)	LDS (for feature smoothing)
[47]	3 audio-video stimuli with 5 trials for each emotion	Raw EEG (0.5-100Hz)	F3 and F4	Multiwavelet Transform (Gernoimo-Hardin-Massopust (GHM), Chui-Lian (CL) and SA4), followed by Euclidian Distance based Features from 3-D Phase Space Reconstruction.	Not performed
[15]	DEAP	θ, α and β	32 electrodes (10-20 system). F3/F4; C3/C4; P3/P4; O3/O4 were used for calculating alpha asymmetry	The power and relative power of θ, α and β bands; the absolute ratio of β power to θ power; and peak-to-peak amplitude; α asymmetry; Shannon, spectral and Kolmogorov entropy; C0-complexity; Skewness, Kurtosis and Variance; The 3 Hjorth parameters	Entropy gains and ratios
[37]	DEAP	θ, α, β and γ	32 electrodes (10-20 system)	3 types of connectivity: Pearson Correlation Connectivity, Phase Coherence Connectivity, Mutual Information	Fisher Linear Discriminant
[48]	MAHNOB	θ, slow α, α, β and γ	32 electrodes (10-20 system)	Logarithmic PSD and Spectral Asymmetry	ANOVA
[17]	SEED	δ, θ, α, β, and γ	62 electrodes (10-10 system)	STFT followed by differential entropy (DE), differential asymmetry (DASM) and rational asymmetry (RASM)	Linear Dynamic Systems (LDS) approach
[49]	DEAP	γ	12 EEG channels (10-20 system)	Wavelet decomposition, followed by Shannon Entropy, Cross-correlation, and Auto-regressive modelling. **(1 feature)**	Not performed
[12]	DEAP	θ, α, β and γ	14 electrodes (10-20 system)	STFT, Hilbert-Huang Spectrum (HHS), Higher Order Crossing (HOC).	mRMR

Ref	Emotion Elicitation	Brain Waves	Channels	Feature extraction	Feature Selection
[13]	DEAP	θ, low α, α, β and γ	32 electrodes (later reduced to 14)	Statistical features (S), band power (BP) for different frequencies, Hjorth parameters (HP) and fractal dimension (FD).	mRMR
[20]	IAPS	δ, θ, α, β e γ	14 electrodes (10-20 system)	6 statistical features, spectral logarithmic PSD, ERP (event related potential). The former was divided into *early, middle e late*	Not performed
[33]	MAHNOB	θ, α, β e γ	32 electrodes (10-20 system)	Logarithmic PSD	Not performed
[50]	DEAP	β	F3 and C4 (since they showed the most "evident" emotional response)	EMD followed by Sample Entropy (SampEn). IMF's with different sizes were used and only those with the highest variance.	Not performed
[31]	SEED (SJTU Emotion EEG Dataset). Famous Chinese films. DEAP	δ, θ, α, β e γ	62 electrodes (10-10 system)	Power spectral density (PSD), differential entropy (DE), differential asymmetry (DASM), rational asymmetry (RASM), asymmetry (ASM) and differential caudality (DCAU). Linear dynamic systems (LDS) for feature smoothing.	PCA and mRMR
[26]	Ten odours were used in the study: rose water, lavender, jasmine, chocolate powder, mint, valerian, garlic powder, anise, cooked cauliflower, and baby shampoo. Selected based on previous studies	θ, α, low β, middle β, high β, and low γ	19 electrodes, taken from 256	PSD, Wasserstein distance and Normalized Length Density (NLD) (EEG features)	Not performed
[51]	MAHNOB-HCI	Raw EEG (1-32Hz)	32 electrodes (10-20 system)	PSD in the 1-32 Hz band. Power in 2 Hz, 4 Hz, 8 Hz, and 16 Hz. Spectral power difference (SPD) between homologue electrodes	ANOVA (p-value < 0.2)
[52]	Thirty 1-minute audio music tracks of five genres namely metal, electronic, rock, hip-hop and rap	Raw EEG (1-50 Hz)	FP1	**Time domain:** Latency to Amplitude Ratio, Peak to Peak Signal Value, Peak to Peak Time Window, Peak to Peak Slope, Signal Power, Mean Value of Signal, Kurtosis, Mobility, Complexity. **Frequency domain:** PSD, Band Power. **Wavelet domain features:** Entropy and Energy	Not performed

Ref	Emotion Elicitation	Brain Waves	Channels	Feature extraction	Feature Selection
[53]	16 pop songs chosen by each subject. 8 songs they were familiar with and 8 songs they were unfamiliar with	δ, θ, α, β e γ	12 electrodes (10-20 system)	Fractal dimensions (FD) and Power Spectrum Density (PSD) + Asymmetries	Not performed
[54]	IADS	θ, α and β	FC5, F4, F7, AF3, and T7 (based on previous works)	6 statistics: mean, standard deviation, mean of absolute values of the first differences, mean of absolute values of the first differences of normalized EEG, mean of absolute values of the second differences, mean of the absolute values of the second differences of the normalized EEG. PSD and HOC	ICC (intra class correlation)
[35]	DEAP	θ, α, β and γ	32 electrodes (10-20 system)	PSD for each band and normalization	ReliefF based selection
[55]	6 excerpts of classical music and 2 excerpts of Iranian music (60s)	θ, α, β and γ	14 electrodes (10-20 system)	MVAR (multivariate autoregressive models) to estimate 6 types of connectivity	Not performed
[14]	DEAP	θ, α, β, γ and noise (64-128Hz)	F3– F4, F7–F8, FC1–FC2, FC5–FC6 and FP1–FP2 (dipoles)	DWT (discrete wavelet transform). Then they calculated the energy and entropy for each band (decomposition level). This was done in 2 and 4s epoch	Not performed
[56]	DEAP	4-45Hz EEG free of artefacts	32 electrodes (10-20 system)	EMD. Then first difference of time series, the first difference of phase, and the normalized energy	Features were tested one at a time. The channels were selected later using Fisher distance
[57]	IAPS	δ, θ, α, β and γ	14 electrodes (10-20 system)	The three Hjorth parameters: activity, mobility e complexity	one-way ANOVA method for a p-value < 0.05
[58]	15 advertisement clips	0-64 Hz EEG free of artefacts	14 electrodes (10-20 system)	Average signal value for each electrode	Variable importance as per IncMSE and variable importance as per IncNodePurity (electrode selection)
[59]	DEAP and MAHNOB	θ, slow α, α, β and γ	32 electrodes (10-20 system)	Theta, slow alpha, alpha, beta, and gamma spectral power for each electrode. The spectral power asymmetry between 14 pairs of electrodes in the four bands of alpha, beta, theta and gamma.	RFE (recurssive feature elimination) and MFE (margin-maximizing feature elimination). Later compared with PCA.
[60]	DEAP + a dataset composed of 16 music excerpts	δ, θ, α, β e γ	12 electrodes (10-20 system)	PSD and FD (Higuchi's method)	Not performed

Ref	Emotion Elicitation	Brain Waves	Channels	Feature extraction	Feature Selection
[61]	DEAP	Raw EEG	32 electrodes (10-20 system)	PSD used in different sized temporal windows. Then MFI (multidimensional feature image) is used	Not performed
[62]	Four music episodes were selected as stimuli; each one was presented for 2 minutes.	θ, α, β e γ	Fp1, Fp2, F7 and F8 (according to previous study)	Mean, Variance, Power, Standard Deviation, 1st difference, normalized 1st difference 2nd difference, Normalized 2nd difference, Hjorth parameters, HOC, Nonlinear Energy. STFT, PSD. Root mean square (RMS), logarithm and the absolute of Wavelet decomposition. Magnitude Squared Coherence Estimate (MSCE)- Differential Asymmetry- Rational Asymmetry. Maximum Lyapunov Exponent. Shannon, Renyi, Tsallis, and Approximation Entropy	mRMR
[34]	DEAP Dataset	4-45Hz EEG free of artefacts	18 channels (eight left, eight right and two central channels in the frontal lobes)	EMD. After normalizing the IMFs, the following were calculated: IMF power, power ratio, PSD, the 3 Hjorth parameters, Entropy, asymmetric power difference, asymmetric power ratios, asymmetric relative entropy, asymmetric correlation, asymmetric coherence	Independent component analysis (ICA)
[36]	DEAP dataset	β	F3, F4, C3, C4 (later reduced to only F3-C4 based on correlation)	EMD followed by autoregressive models (AR). Tested different number of AR coefficients, and only using the 4 first IMFs (higher variance)	Not performed
[63]	DREAMER (18 film clips)	θ, α e β	14 electrodes (10-20 system)	Logarithmic PSD	Not performed

Ref	Emotion Elicitation	Brain Waves	Channels	Feature extraction	Feature Selection
[64]	MAHNOB (15 subjects used) + DEAP (15 subjects used) + a new dataset comprised of EEG taken from 13 subjects (later reduced to 11) that watched video clips	θ, α, β e γ	32 electrodes for DEAP and MANHOB, and 5 electrodes for the newly constructed dataset (AF3, AF4, T7, T8, Pz)	**Time domain**- Variance Quartile 1, quartile median, quartile 3, quartile 4 Mean Minimum, Maximum Standard deviation, Normalized 1st difference Normalized second difference, Non-stationary index HOC, Root mean square (RMS), **Frequency domain**- Power spectrum density (PSD) from Gamma, Theta, Alpha, Beta, **Time–frequency domain-** Power of Discrete Wavelet Transform (DWT) from Gamma, Theta, Alpha, Beta, RMS of DWT from Gamma, Theta, Alpha, Beta Log (REE) of DWT from Gamma, Theta, Mean Power, 1st difference 2nd difference, 3 Hjorth parameters FD (Higuchi algorithm), variance	Ant colony optimization (ACO), Simulated annealing (SA), Genetic algorithm (GA), Particle swarm optimization (PSO), Differential evolution (DE)
[65]	DEAP and SEED	θ, α, β e γ	32 and 62 electrodes (DEAP and SEED, respectively)	**Time-frequency domain features-** Peak-Peak Mean. Mean Square Value. Variance. Hjorth Parameter: Activity. Hjorth Parameter: Mobility. Hjorth Parameter: Complexity. Maximum Power Spectral Frequency. Maximum Power **Non-linear dynamical system features-** Approximate Entropy. C0 Complexity. Correlation Dimension. Kolmogorov Entropy. Lyapunov Exponent. Permutation Entropy. Singular Entropy. Shannon Entropy. Spectral Entropy	Chi-Squared-Based Feature Selection (χ2), Mutual Information-Based Feature Selection (MI), ANOVA, Recursive Feature Elimination (RFE), L1-Norm Penalty-Based Feature Selection (L1)
[18]	AMIGOS. Two experimental settings. In the first one, 40 participants watched 16 short emotional videos. In the second one, they watched 4 long videos, some of them alone and the rest in groups.	θ, slow α, α, β e γ	14 electrodes (10-20 system)	Logarithmic PSD using Welch method	Fisher Linear Discriminant

Ref	Emotion Elicitation	Brain Waves	Channels	Feature extraction	Feature Selection
[66]	DEAP	4-45Hz EEG free of artefacts	22 electrodes (10-20 system)	Quadratic time-frequency distributions (QTFDs) using Choi-Williams distribution. T**ime-domain features**: mean (μ), variance (σ), skewness (γ), kurtosis (κ), sum of the logarithmic amplitudes (SLA), median absolute deviation (MAD), root mean square value (RMS) and inter-quartile range (IQR). **Frequency-domain features**: flatness (FLS), flux (FLX), spectral roll-off (SRO), normalized Renyi entropy (NRE) and energy concentration (EC)	mRMR
[67]	16 Chinese film clips: 2 per emotion	δ, θ, α, β e γ	AF3-AF4, F3-F4, F7-F8, FC5-FC6, T7-T8, P7- P8, and O1-O2. The electrodes were also analysed individually	STFT with subtracted baseline	LDA (linear discriminant analysis)
[68]	MAHNOB-HCI	β	32 later reduced to only 2: F3 and C4.	EMD method to decompose the signal. Then, SampEn is applied on the first 4 components.	Black hole algorithm
[69]	DEAP	θ, α, β and γ	32 electrodes (10-20 system)	Spectral features based on Fourier spectrum plus amplitude modulation features based on the Hilbert spectrum: AMI (amplitude modulation interaction), AMC (amplitude modulation coherence), AME (amplitude modulation energy), PAC (phase-amplitude coupling).	mRMR followed by ANOVA
[70]	28 videos describing seven different emotion states (joy, funny, anger, disgust, fear, sad and neutrality)	δ, θ, α, β and γ	62 electrodes (10-10 system)	Hjorth parameters, HOC, logarithmic PSD (per band), HHS	Not performed
[71]	MAHNOB-HCI	Low α, high α, β and γ	14 electrodes (10-20 system)	PSD using Welch method	Not performed
[72]	DEAP	0.5-47 Hz EEG free of artefacts	15 electrodes, associated with Brodmann regions of audio-visual stimuli processes	Wavelet decomposition. Each level represented one frequency band (δ, θ, α, β and γ). The coefficients were used directly as features	Not performed

Table A1- All the works consulted so far, organized according to the feature extraction methods used.

Ref	Participants	Gender Specific	Classes	Classification Method	Best Accuracies
[39]	26 subjects	No	4 quadrants: High/Low Valence; High/Low Arousal (joy, angry, sadness and pleasure)	SVM	92.73% (individual score), when using all frequency bands
[40]	26 subjects	No	4 quadrants: High/Low Valence; High/Low Arousal (joy, angry, sadness and pleasure)	SVM. 3 different methods: *all together*, *one against one*, *model based*	94.86 for valence
[22]	26 participants, 16 males and10 females, mean age=24.40±2.53	No	4 quadrants: High/Low Valence; High/Low Arousal (joy, angry, sadness and pleasure)	Multilayer Perceptron (MLP) and SVM	82.29%
[25]	16 participants (9 males and 7 females)	No	Happiness, surprise, anger, fear, disgust, and sadness	Quadratic Discriminant Analysis (QDA), k-NN, Mahalanobis Distance (MD), SVM. Leave-one-out (LOO) and cross-validation	85% (for SVM and combined features)
[41]	3 females and 17 males in the age group of 21-39	No	disgust, happy, surprise, fear and neutral	KNN (k=2-6) and LDA	91.67% for disgust, 81.67% for happy and surprise, 81.25% for fear and 93.75% for neutral
[38]	9 male subjects (23-27 years old)	No	Positive/negative valence	LDA	55.5% (individual score) using CSP
[21]	Set1-(10 participants- 8 males and 2 females); Set2-(12 participants-9 males and 3 females)	No	6 classes: fear, frustrated, sad, happy, pleasant and satisfied	Not stated	Not stated
[19]	16 participants (9 males and 7 females)	No	High/Low Valence; High/Low Arousal	SVM with a five-order polynomial kernel, after having tested other classifiers	≈60%. Hoc was the best method and the group of large AsI generated the best results
[30]	Set1-(10 participants- 8 males and 2 females); Set2-(12 participants-9 males and 3 females)	No	negative high aroused (fear), positive high aroused, (happy), negative low aroused (sad), and positive low aroused (pleasant)	SVM	≈100% (individual score)
[11]	32 participants (16 males and 16 females; mean age=26.9)	No	High/Low Valence; High/Low Arousal; High/Low Liking	naive Bayes	62% for Arousal; 57.6% for Valence (EEG)
[16]	28 participants (12 males and 16 females)	No	Low arousal, medium aroused, activated, unpleasant, neutral valence, pleasant	SVM with RBF kernel	67.7% and 76.1% for arousal and valence respectively (EEG and Eye Gaze data fused)
[42]	16 participants	No	4 quadrantes: High/Low Valence; High/Low Arousal	SVM	74.28%
[43]	6 healthy subjects (3 males and 3 females)	No	4 quadrantes: High/Low Valence; High/Low Arousal	Linear Discriminant Analysis (LDA) and Support Vector Machines (SVM linear and RBF)	83.35%, and 86.33% (individual scores for valence and arousal respectively)
[44]	3 females and 17 males in the age group of 21-39	No	disgust, happy, surprise, fear and neutral	KNN and PNN (probabilistic neural networks)	91.33%, for SE and beta frequency and for both classifiers

Ref	Participants	Gender Specific	Classes	Classification Method	Best Accuracies
[24]	16 young adult volunteers (mixed gender), later reduced to 12	No	12 emotions (awe, jealousy, contentment, frustration, relief, joy, fear, love, anger, sadness, happiness, grief)	Logistic Regression	71.3±14.9%
[32]	32 participants (16 males and 16 females; mean age=26.9)	No	6 classes: 3 valence classes and 3 arousal classes	DLN (deep learning), in 4 different sets: DLN100 (100 nodes in each layer); DLN50; PCA+DLN50; PCA+CSA (covariate shift adaptation) +DLN50. These were later compared with SVM and naive Bayes	53.42±9.64% (Valence) and 52.03±9.74% (Arousal)- PCA+CSA+DLN50 was the best set
[23]	6 participants (3 males and 3 females)	No	2 classes (positive and negative emotion)	SVM with 3 different kernels (linear, polynomial and RBF)	82.38% with SVM and using power spectrum features. The other features generated accuracies around 70%, but with a much lower error
[45]	40 healthy, right-handed students from National Cheng Kung University (21 males and 19 females)	Yes	Positive, neutral negative	QDA	mean=82% using PSI
[46]	5 subjects participated (2 females and 3 males whose ages range from 22-24)	No	positive, neutral and negative	SVM linear	71.77% using DE
[47]	8 healthy subjects (4 males and 4 females; age 20-35 years)	No	happy, neutral, sadness, and fear	LS-SVM (least squares) (RBF, Morlet and Mexican Hat)	98.33 for fear
[15]	32 participants (16 males and 16 females; mean age=26.9)	Yes	High/Low Valence; High/Low Arousal	MLP, SVM, K-NN, Decision Tree (C4.5)	69% using C4.5 classifier
[37]	32 participants (16 males and 16 females; mean age=26.9)	No	High/Low Valence; High/Low Arousal	SVM	76%
[48]	30 participants (13 males and 17 females). Only the data of 24 participants was used	No	3 classes de valence (unpleasant, neutral, pleasant) e 3 classes de arousal (calm, medium, activated)	SVM classifier with RBF kernel	76.4% for arousal and 68.5% for valence when using DLF (decision level fusion) with EEG and eye gaze
[17]	(15 participants 7 males and 8 females)	No	Positive, neutral, and negative	k-NN, logistic regression (LR), SVM, DBN (deep belief networks)	86.08±8.34 (DBN) Beta and gamma band features generated the highest accuracies
[49]	32 participants (16 males and 16 females; mean age=26.9)	No	four levels of emotions, namely excitation, happiness, sadness and hatred.	ML-SVM (multiclass SVM)	96.52% for sadness classification
[12]	32 participants (16 males and 16 females; mean	No	Anger, Surprise, Other	Random Forest (RF) e Support Vector Machine	40-60%. RF was considerably more accurate

Ref	Participants	Gender Specific	Classes	Classification Method	Best Accuracies
	age=26.9)			(SVM)	
[13]	32 participants (16 males and 16 females; mean age=26.9)	No	High/Low Valence; High/Low Arousal	SVM	62.39%
[20]	21 participants (9 males and 12 females)	No	Happy, scared, calmed, sad (one emotion for each quadrant in V/A space)	K-NN e SVM	≈58% using the 3 LPPs (late positive potential), and SVM
[33]	28 participants (12 males and 16 females).	No	There isn't emotion classification. Values of V/A are being estimated along with correlations and errors	Multi-linear regression (MLR), SVR, CCRF (Continuous Conditional Random Fields), and LSTM-RNN (Long Short-Term Memory Neural Networks)	**EEG-** ρ=0.26±0.49; using CCRF **EEG+Face (DLF)-** ρ=0.45±0.35; using LSTM-RNN
[50]	32 participants (16 males and 16 females; mean age=26.9)	No	High/Low Valence; High/Low Arousal	SVM with Gaussian kernel	93.20%
[31]	(15 participants 7 males and 8 females)	No	High/Low Valence; High/Low Arousal	SVM, KNN, Logistic Regression (LR) e discriminative Graph regularized Extreme Learning Machine (GELM)	91.07% (SEED)
[26]	25 subjects (9 females and 16 males) (age 24± 4.6 years old)	No (after having performed a Kruskal-Wallis test)	3: unpleasant, pleasant and neutral	Linear discriminant analysis (LDA)	99% for unpleasant. The recognition accuracy for the other emotions was close to 50%
[51]	28 participants (12 males and 16 females).	No	High/Low Valence; High/Low Arousal	SVM and KNN (K=3)	62.13% for valence and 61.80% for arousal (using EEG features)
[52]	30 participants (15 males and 15 females) (15-50 years old)	No	Happy, love, sad, anger	SVM, KNN e MLP	98% accuracy for the middle-aged group using KNN
[53]	Fifteen males between 22 and 30 years of age (mean = 25.52, SD = 2.14) participated in the experiments	No	High/Low Valence; High/Low Arousal	support vector machine (SVM), MLP, and C4.5	87.2% for valence classification, and using FD features
[54]	5 subjects, 4 males and 1 female, aged 24–28	No	pleasant, happy, angry and frightened	SVM polynomial kernel	93.37%
[35]	32 participants (16 males and 16 females; mean age=26.9)	No	High/Low Valence; High/Low Arousal (joy, fear, sadness, relaxation)	SVM (RBF)	62.59%
[55]	19 participants (11 females and 8 males)	No	High/Low Valence; High/Low Arousal in three modalities: joyful vs melancholic; joyful vs neutral; familiar vs unfamiliar	SVM (8-fold cross validation)	93.7% ± 1.06%, 80.43% ± 1.74%, 83.04% ± 1.47 (joyful vs. neutral, joyful vs. melancholic, and familiar vs. unfamiliar, respectively)

Ref	Participants	Gender Specific	Classes	Classification Method	Best Accuracies
[14]	32 participants (16 males and 16 females; mean age=26.9)	No	High/Low Valence; High/Low Arousal	SVM and K-NN	86.75%, using KNN, 4s epochs, and after using the data taken from 5 pairs of channels
[56]	32 participants (16 males and 16 females; mean age=26.9)	No	High/Low Valence; High/Low Arousal	SVM com RBF kernel	72% for arousal e 69% para valence
[57]	21 subjects (12 male and 9 female) (12-14 years old)	No	4 quadrantes: High/Low Valence; High/Low Arousal	LDA, KNN, SVM, naïve-Bayes, random- forest, deep-learning and ensemble methods.	76.6% for voting (ensemble method)
[58]	25 participants (Mean age= 28.1, Standard Deviation = 4.98)	No	Positivity and negativity score (0-5). The "self-assessment" was generated by the comments to the clips on YouTube	3 regressors: linear, decision tree and random forest	R^2=0.68
[59]	32 participants (16 males and 16 females; mean age=26.9)- DEAP; 28 participants (12 males and 16 females)- MAHNOB	No	High/Low Valence; High/Low Arousal	SVM with RBF kernel	74.82%
[60]	15 healthy subjects between 22 and 30 years of age (mean = 25.52, SD = 2.14)	No	High/Low Valence; High/Low Arousal	SVM (PUK), MLP with one hidden layer, and C4.5.	87.80%
[61]	32 participants (16 males and 16 females; mean age=26.9)	No	4 quadrantes: High/Low Valence; High/Low Arousal	KNN, SVM (RBF), RF (random forest), CNN (convolutional neural networks) + RNN (recurrent neural networks), CLRNN	90.54% (individual score) using CLRNN
[62]	16 participants (8 males and 8 females) aged between 19 and 37 (mean age 24.9 ± 9.08 years old), selected from four ethnic groups, namely, Kurd, Turk, Lur, and Gilak (each group consisting of four participants)	No	Level of joy (positive vs negative)	SVM (RBF) with different metrics: Euclidean, City-Block (Manhattan), Minkowsky, Chebyshev, Cosine, Correlation, Spearman, DTW, Mahalanobis	83.30%
[34]	32 participants (16 males and 16 females; mean age=26.9)	No	High/Low Valence; High/Low Arousal	KNN and Artificial Neural Networks (ANN). *leave one out* (LOO) method	75% (para ANN)
[36]	32 participants (16 males and 16 females; mean age=26.9)	No	High/Low Valence; High/Low Arousal	SVM with Gaussian kernel. 10-fold cross validation. 4 classifiers were made, one for each combination of V/A	92.04% (HAHV/HALV)
[63]	23 participants (14 males and 9 females)	No	low/high arousal (calm/excited), low/high valence (unpleasant/pleasant), and low/high dominance (without control/empowered)	3 SVM classifiers with RBF kernel. One for valence, one for arousal, and another for dominance	≈62% para valence e arousal

Ref	Participants	Gender Specific	Classes	Classification Method	Best Accuracies
[64]	13 subjects, aged between 20 and 38 years old, while they watched video clips (nova dataset).	No	High/Low Valence; High/Low Arousal	Probabilistic neural network (PNN)	≈97%, in MANHOB dataset using GA (far greater accuracy than the other datasets). Time domain features were the top chosen by the selector along with frontal and central channels
[65]	32 and 15 (DEAP and SEED, respectively)	No	2 classes (positive emotions and negative emotions)	Linear SVM (leave-one-out)	Highest mean recognition accuracies of 59.06% on the DEAP dataset and 83.33% on the SEED dataset. L1 selector generated the highest accuracies. In the SEED dataset, temporal lobes generated the best results
[18]	40 (27 males and 13 females) in short videos and 37 participants in long videos	No	High/Low Valence; High/Low Arousal	Gaussian naive Bayes	Valence F1-score =0.576; Arousal F1-score=0.592 (averages)
[66]	32 participants (16 males and 16 females; mean age=26.9)	No	1D-2CLS (positive/negative); 1D-3CLS (positive/negative/neutral); 2D-4CLS (High/Low valence/arousal); 2D-5CLS (High/Low valence/arousal+neutral)	SVM with RBF kernel	92.7% (score individual) para 1D-2CLS e Arousal.
[67]	30 participants (age range: 19-26, mean = 23, SD = 1.73 years)	No	joy, amusement, tenderness, sadness, anger, fear, disgust and neutral	SVM (several models were created to encompass every possible combination of emotion pairs)	92.26% in recognising high-arousal and valence emotions from neutrality.
[68]	28 participants (12 males and 16 females).	No	4 quadrantes: High/Low Valence; High/Low Arousal	SVM	93%
[69]	32 participants (16 males and 16 females; mean age=26.9)	No	High/Low Valence; High/Low Arousal	SVM with RBF kernel (chosen after pilot tests)	60.4% and 58.3% for valence and arousal respectively
[70]	23 participants (mean age=20)	No	joy, funny, anger, sad, disgust, fear and neutrality. The classifications are performed between one positive emotion, one negative and the neutral state	A-LSTM, SVM e KNN.LSTM+X, LSTM+H, LSTM+C (3 different modalities with LSTM)	83.07%
[71]	28 participants (12 males and 16 females).	No	3 classes of valence, 3 classes of arousal and 3 classes of emotion	KNN (Euclidian distance). 70/30 ratio for train/test	96.341%, 96.341%, 96.222%, for emotion, valence and arousal classification, respectively

Ref	Participants	Gender Specific	Classes	Classification Method	Best Accuracies
[72]	32 participants (16 males and 16 females; mean age=26.9)	No	4 quadrantes: High/Low Valence; High/Low Arousal. Two class identification, 3 class identification and 4 class identification schemes	NN and SVM (RBF, polynomial and multi-layer perceptron)	90.2%, 84.2% and 80.9%, for the 2, 3 and 4 class identification scheme (respectively), using NN. 88.55%, 83.26%, 81.97% for the 2, 3 and 4 class identification scheme (respectively), using SVM with the polynomial kernel

Table A2- All the works consulted so far, organized according to their emotion classification methods. Some of the methods used were written with their initials only. For clarification one can see their meaning above the table.

Appendix B

Electrode Channel	Band	Estimator	Arousal			Valence		
			PCC	MAE	RMSE	PCC	MAE	RMSE
AF3	Gamma	KNN	0.084	0.232	0.301	0.116	0.296	0.377
		RF	0.283	0.176	0.214	0.314	0.225	0.268
	Beta	KNN	0.094	0.230	0.300	0.130	0.291	0.375
		RF	0.284	0.176	0.214	0.333	0.223	0.266
	Alpha	KNN	0.085	0.234	0.302	0.098	0.298	0.380
		RF	0.256	0.179	0.215	0.297	0.229	0.269
AF4	Gamma	KNN	0.078	0.233	0.302	0.110	0.295	0.378
		RF	0.297	0.175	0.213	0.296	0.227	0.269
	Beta	KNN	0.087	0.231	0.301	0.126	0.292	0.374
		RF	0.300	0.175	0.213	0.320	0.225	0.267
	Alpha	KNN	0.063	0.237	0.305	0.101	0.298	0.379
		RF	0.239	0.180	0.216	0.287	0.230	0.270
F3	Gamma	KNN	0.076	0.233	0.302	0.111	0.294	0.376
		RF	0.299	0.175	0.213	0.317	0.226	0.267
	Beta	KNN	0.083	0.232	0.301	0.114	0.293	0.376
		RF	0.301	0.175	0.212	0.312	0.226	0.268
	Alpha	KNN	0.065	0.238	0.305	0.094	0.300	0.381
		RF	0.255	0.179	0.215	0.274	0.231	0.271
F4	Gamma	KNN	0.084	0.232	0.302	0.106	0.296	0.379
		RF	0.289	0.176	0.213	0.305	0.226	0.268
	Beta	KNN	0.100	0.230	0.300	0.125	0.290	0.375
		RF	0.302	0.175	0.212	0.323	0.224	0.267
	Alpha	KNN	0.089	0.233	0.302	0.112	0.294	0.377
		RF	0.275	0.177	0.214	0.276	0.230	0.271
F7	Gamma	KNN	0.093	0.230	0.299	0.111	0.294	0.377
		RF	0.298	0.174	0.213	0.285	0.228	0.270
	Beta	KNN	0.099	0.229	0.300	0.128	0.290	0.374
		RF	0.299	0.174	0.213	0.305	0.227	0.268
	Alpha	KNN	0.087	0.233	0.301	0.084	0.302	0.383
		RF	0.250	0.179	0.216	0.262	0.232	0.272
F8	Gamma	KNN	0.088	0.231	0.300	0.103	0.296	0.379
		RF	0.274	0.177	0.214	0.310	0.226	0.268
	Beta	KNN	0.106	0.228	0.298	0.118	0.292	0.375
		RF	0.279	0.176	0.214	0.332	0.223	0.266
	Alpha	KNN	0.088	0.235	0.301	0.099	0.299	0.380
		RF	0.275	0.178	0.214	0.292	0.229	0.270
FC5	Gamma	KNN	0.087	0.232	0.301	0.098	0.300	0.381
		RF	0.266	0.177	0.215	0.292	0.228	0.270
	Beta	KNN	0.093	0.230	0.301	0.115	0.293	0.377
		RF	0.282	0.176	0.214	0.301	0.227	0.269
	Alpha	KNN	0.077	0.235	0.303	0.092	0.300	0.381
		RF	0.264	0.232	0.272	0.246	0.180	0.216

FC6	**Gamma**	**KNN**	0.093	0.231	0.301	0.117	0.294	0.376
		RF	0.253	0.178	0.216	0.307	0.226	0.268
	Beta	**KNN**	0.084	0.232	0.302	0.115	0.295	0.377
		RF	0.261	0.178	0.215	0.321	0.225	0.267
	Alpha	**KNN**	0.079	0.235	0.303	0.102	0.299	0.378
		RF	0.219	0.181	0.218	0.274	0.231	0.271
O1	**Gamma**	**KNN**	0.078	0.234	0.303	0.101	0.297	0.379
		RF	0.280	0.176	0.214	0.278	0.230	0.271
	Beta	**KNN**	0.103	0.228	0.298	0.118	0.292	0.376
		RF	0.306	0.174	0.212	0.284	0.229	0.270
	Alpha	**KNN**	0.077	0.235	0.303	0.103	0.299	0.379
		RF	0.280	0.177	0.214	0.265	0.231	0.272
O2	**Gamma**	**KNN**	0.075	0.234	0.303	0.096	0.299	0.381
		RF	0.293	0.175	0.213	0.278	0.229	0.271
	Beta	**KNN**	0.095	0.229	0.300	0.116	0.293	0.375
		RF	0.303	0.174	0.212	0.289	0.229	0.270
	Alpha	**KNN**	0.078	0.235	0.303	0.109	0.296	0.378
		RF	0.269	0.177	0.215	0.262	0.231	0.272
P7	**Gamma**	**KNN**	0.101	0.228	0.299	0.120	0.292	0.377
		RF	0.306	0.174	0.212	0.317	0.225	0.267
	Beta	**KNN**	0.099	0.230	0.299	0.125	0.291	0.375
		RF	0.309	0.174	0.212	0.307	0.226	0.268
	Alpha	**KNN**	0.096	0.232	0.299	0.097	0.298	0.380
		RF	0.263	0.178	0.215	0.274	0.231	0.271
P8	**Gamma**	**KNN**	0.099	0.229	0.299	0.109	0.294	0.378
		RF	0.311	0.174	0.212	0.310	0.225	0.268
	Beta	**KNN**	0.111	0.228	0.298	0.118	0.291	0.376
		RF	0.318	0.173	0.211	0.328	0.223	0.266
	Alpha	**KNN**	0.090	0.233	0.301	0.111	0.296	0.378
		RF	0.312	0.175	0.212	0.291	0.228	0.270
T7	**Gamma**	**KNN**	0.095	0.231	0.300	0.121	0.292	0.375
		RF	0.290	0.175	0.213	0.308	0.225	0.268
	Beta	**KNN**	0.098	0.230	0.300	0.138	0.288	0.371
		RF	0.300	0.174	0.212	0.326	0.223	0.266
	Alpha	**KNN**	0.090	0.234	0.301	0.095	0.299	0.381
		RF	0.268	0.178	0.215	0.279	0.230	0.271
T8	**Gamma**	**KNN**	0.089	0.232	0.301	0.111	0.295	0.377
		RF	0.303	0.175	0.212	0.330	0.224	0.266
	Beta	**KNN**	0.108	0.227	0.297	0.100	0.295	0.378
		RF	0.314	0.174	0.211	0.344	0.222	0.265
	Alpha	**KNN**	0.071	0.235	0.302	0.090	0.299	0.381
		RF	0.266	0.178	0.215	0.294	0.229	0.269

Table B1- V/A estimation results obtained when comparing each electrode channel individually, using AMIGOS dataset.

Band	Feature	Estimator	Arousal			Valence		
			PCC	MAE	RMSE	PCC	MAE	RMSE
Gamma	H1	KNN	0.681	0.086	0.178	0.658	0.117	0.233
		RF	0.741	0.121	0.157	0.723	0.164	0.204
	H2	KNN	0.206	0.201	0.280	0.224	0.255	0.351
		RF	0.320	0.175	0.211	0.339	0.226	0.266
	H3	KNN	0.110	0.226	0.297	0.120	0.288	0.374
		RF	0.245	0.180	0.216	0.279	0.231	0.271
	SE	KNN	0.081	0.233	0.302	0.109	0.296	0.377
		RF	0.219	0.181	0.217	0.259	0.233	0.272
	WP	KNN	0.679	0.086	0.179	0.658	0.116	0.233
		RF	0.761	0.115	0.151	0.746	0.156	0.197
	WE	KNN	0.236	0.194	0.274	0.234	0.252	0.349
		RF	0.353	0.173	0.209	0.352	0.224	0.264
	IMFP	KNN	0.472	0.141	0.230	0.440	0.188	0.298
		RF	0.734	0.121	0.157	0.705	0.168	0.208
	IMFE	KNN	0.311	0.176	0.260	0.297	0.231	0.335
		RF	0.401	0.169	0.205	0.380	0.222	0.262
Beta	H1	KNN	0.635	0.098	0.190	0.584	0.139	0.256
		RF	0.753	0.116	0.152	0.722	0.163	0.203
	H2	KNN	0.300	0.179	0.264	0.316	0.227	0.329
		RF	0.398	0.168	0.205	0.395	0.218	0.260
	H3	KNN	0.194	0.206	0.282	0.206	0.264	0.355
		RF	0.339	0.173	0.210	0.358	0.222	0.263
	SE	KNN	0.104	0.227	0.297	0.118	0.291	0.374
		RF	0.245	0.180	0.216	0.276	0.231	0.271
	WP	KNN	0.657	0.093	0.185	0.614	0.130	0.248
		RF	0.770	0.112	0.148	0.740	0.158	0.198
	WE	KNN	0.332	0.172	0.258	0.316	0.225	0.330
		RF	0.434	0.165	0.202	0.399	0.219	0.260
	IMFP	KNN	0.296	0.182	0.265	0.250	0.248	0.346
		RF	0.559	0.152	0.188	0.495	0.210	0.248
	IMFE	KNN	0.240	0.196	0.274	0.227	0.256	0.352
		RF	0.344	0.174	0.209	0.331	0.227	0.267
Alpha	H1	KNN	0.393	0.156	0.245	0.345	0.214	0.323
		RF	0.603	0.145	0.181	0.534	0.205	0.243
	H2	KNN	0.256	0.189	0.271	0.255	0.246	0.344
		RF	0.336	0.174	0.210	0.338	0.228	0.266
	H3	KNN	0.112	0.224	0.296	0.131	0.287	0.372
		RF	0.249	0.180	0.216	0.268	0.233	0.272
	SE	KNN	0.088	0.231	0.299	0.108	0.293	0.376
		RF	0.221	0.181	0.217	0.261	0.233	0.272
	WP	KNN	0.431	0.148	0.237	0.383	0.204	0.314
		RF	0.645	0.139	0.175	0.574	0.197	0.236
	WE	KNN	0.352	0.169	0.253	0.323	0.224	0.329
		RF	0.429	0.167	0.202	0.395	0.221	0.260
	IMFP	KNN	0.335	0.172	0.254	0.318	0.180	0.260
		RF	0.590	0.157	0.191	0.558	0.160	0.195
	IMFE	KNN	0.268	0.192	0.276	0.246	0.190	0.278
		RF	0.453	0.178	0.251	0.436	0.180	0.253

Table B2- V/A estimation results obtained when comparing each feature extraction method individually, using AMIGOS dataset.

Appendix C

Electrode Channel	Band	Estimator	Arousal			Valence		
			PCC	MAE	RMSE	PCC	MAE	RMSE
AF3	Gamma	KNN	0.077	0.265	0.343	0.102	0.280	0.359
		RF	0.309	0.197	0.241	0.321	0.211	0.254
	Beta	KNN	0.097	0.261	0.339	0.105	0.277	0.357
		RF	0.343	0.195	0.238	0.327	0.210	0.253
	Alpha	KNN	0.109	0.260	0.338	0.094	0.282	0.361
		RF	0.264	0.201	0.245	0.238	0.217	0.261
AF4	Gamma	KNN	0.084	0.266	0.343	0.093	0.282	0.362
		RF	0.296	0.198	0.242	0.294	0.212	0.256
	Beta	KNN	0.137	0.254	0.333	0.119	0.274	0.356
		RF	0.348	0.193	0.237	0.314	0.210	0.254
	Alpha	KNN	0.099	0.261	0.339	0.093	0.281	0.361
		RF	0.324	0.196	0.240	0.280	0.213	0.257
F3	Gamma	KNN	0.079	0.266	0.344	0.091	0.281	0.360
		RF	0.298	0.197	0.242	0.406	0.201	0.245
	Beta	KNN	0.100	0.260	0.338	0.100	0.278	0.358
		RF	0.418	0.186	0.230	0.421	0.199	0.243
	Alpha	KNN	0.095	0.262	0.340	0.092	0.280	0.360
		RF	0.330	0.195	0.239	0.297	0.213	0.256
F4	Gamma	KNN	0.074	0.267	0.344	0.094	0.282	0.361
		RF	0.274	0.200	0.244	0.405	0.202	0.245
	Beta	KNN	0.093	0.263	0.342	0.108	0.276	0.358
		RF	0.423	0.186	0.230	0.406	0.202	0.245
	Alpha	KNN	0.084	0.265	0.343	0.095	0.281	0.360
		RF	0.264	0.201	0.245	0.244	0.217	0.260
F7	Gamma	KNN	0.089	0.266	0.343	0.105	0.279	0.358
		RF	0.314	0.196	0.241	0.389	0.203	0.247
	Beta	KNN	0.128	0.256	0.335	0.118	0.274	0.355
		RF	0.411	0.187	0.231	0.397	0.203	0.246
	Alpha	KNN	0.114	0.259	0.337	0.094	0.280	0.360
		RF	0.348	0.193	0.238	0.297	0.213	0.256
F8	Gamma	KNN	0.076	0.266	0.344	0.104	0.278	0.358
		RF	0.288	0.198	0.243	0.415	0.200	0.244
	Beta	KNN	0.098	0.262	0.340	0.107	0.276	0.357
		RF	0.440	0.184	0.227	0.423	0.199	0.243
	Alpha	KNN	0.100	0.261	0.340	0.094	0.280	0.361
		RF	0.341	0.194	0.238	0.296	0.213	0.256
FC5	Gamma	KNN	0.080	0.267	0.344	0.097	0.281	0.360
		RF	0.271	0.199	0.244	0.390	0.203	0.247
	Beta	KNN	0.105	0.259	0.340	0.116	0.275	0.356
		RF	0.416	0.186	0.230	0.401	0.202	0.245
	Alpha	KNN	0.097	0.262	0.340	0.102	0.278	0.359
		RF	0.342	0.193	0.238	0.290	0.213	0.256
FC6	Gamma	KNN	0.080	0.266	0.344	0.110	0.277	0.356

		RF	0.273	0.199	0.244	0.429	0.198	0.242
	Beta	**KNN**	0.103	0.259	0.338	0.126	0.274	0.354
		RF	0.416	0.186	0.230	0.415	0.200	0.244
	Alpha	**KNN**	0.097	0.261	0.339	0.098	0.280	0.360
		RF	0.346	0.193	0.238	0.292	0.213	0.256
FP1	**Gamma**	**KNN**	0.080	0.268	0.343	0.097	0.281	0.360
		RF	0.298	0.198	0.242	0.309	0.212	0.255
	Beta	**KNN**	0.098	0.263	0.340	0.111	0.278	0.357
		RF	0.312	0.196	0.241	0.320	0.211	0.254
	Alpha	**KNN**	0.090	0.265	0.342	0.089	0.282	0.361
		RF	0.236	0.202	0.247	0.250	0.217	0.260
FP2	**Gamma**	**KNN**	0.084	0.266	0.343	0.106	0.279	0.357
		RF	0.302	0.198	0.242	0.323	0.210	0.254
	Beta	**KNN**	0.111	0.260	0.338	0.112	0.276	0.356
		RF	0.342	0.194	0.238	0.339	0.209	0.252
	Alpha	**KNN**	0.096	0.262	0.339	0.100	0.279	0.358
		RF	0.336	0.195	0.239	0.284	0.214	0.257
FC1	**Gamma**	**KNN**	0.087	0.265	0.342	0.106	0.279	0.359
		RF	0.302	0.198	0.242	0.415	0.200	0.244
	Beta	**KNN**	0.121	0.255	0.336	0.134	0.271	0.353
		RF	0.402	0.188	0.232	0.404	0.201	0.245
	Alpha	**KNN**	0.093	0.262	0.341	0.094	0.281	0.362
		RF	0.311	0.197	0.241	0.279	0.213	0.257
FC2	**Gamma**	**KNN**	0.082	0.266	0.343	0.103	0.279	0.358
		RF	0.297	0.198	0.242	0.428	0.198	0.242
	Beta	**KNN**	0.103	0.259	0.339	0.126	0.273	0.353
		RF	0.407	0.188	0.231	0.412	0.200	0.244
	Alpha	**KNN**	0.088	0.263	0.341	0.091	0.281	0.361
		RF	0.330	0.195	0.239	0.287	0.213	0.257
C3	**Gamma**	**KNN**	0.086	0.265	0.341	0.105	0.279	0.359
		RF	0.288	0.198	0.243	0.443	0.197	0.240
	Beta	**KNN**	0.115	0.257	0.336	0.133	0.272	0.353
		RF	0.428	0.185	0.229	0.428	0.199	0.242
	Alpha	**KNN**	0.102	0.261	0.340	0.093	0.280	0.362
		RF	0.251	0.201	0.246	0.248	0.216	0.260
C4	**Gamma**	**KNN**	0.046	0.266	0.343	0.112	0.279	0.357
		RF	0.280	0.198	0.243	0.304	0.212	0.255
	Beta	**KNN**	0.088	0.264	0.343	0.117	0.276	0.357
		RF	0.300	0.197	0.242	0.322	0.211	0.253
	Alpha	**KNN**	0.074	0.266	0.344	0.109	0.278	0.358
		RF	0.253	0.201	0.246	0.254	0.216	0.259
CP1	**Gamma**	**KNN**	0.078	0.267	0.345	0.085	0.282	0.362
		RF	0.285	0.199	0.243	0.289	0.213	0.256
	Beta	**KNN**	0.095	0.262	0.341	0.106	0.277	0.358
		RF	0.309	0.197	0.241	0.308	0.211	0.255
	Alpha	**KNN**	0.123	0.258	0.336	0.106	0.277	0.359
		RF	0.343	0.194	0.238	0.292	0.212	0.256
CP2	**Gamma**	**KNN**	0.076	0.268	0.345	0.101	0.279	0.359
		RF	0.287	0.199	0.243	0.276	0.205	0.248
	Beta	**KNN**	0.104	0.260	0.339	0.108	0.276	0.357
		RF	0.395	0.188	0.233	0.396	0.203	0.246
	Alpha	**KNN**	0.098	0.262	0.341	0.104	0.279	0.358
		RF	0.343	0.194	0.238	0.291	0.213	0.256
CP5	**Gamma**	**KNN**	0.085	0.265	0.343	0.101	0.280	0.360
		RF	0.278	0.199	0.243	0.298	0.212	0.256
	Beta	**KNN**	0.115	0.258	0.337	0.122	0.274	0.355
		RF	0.313	0.197	0.241	0.317	0.210	0.254

	Alpha	**KNN**	0.128	0.256	0.334	0.094	0.280	0.360
		RF	0.347	0.194	0.238	0.282	0.213	0.257
CP6	**Gamma**	**KNN**	0.092	0.264	0.341	0.110	0.278	0.358
		RF	0.298	0.197	0.242	0.396	0.202	0.246
	Beta	**KNN**	0.109	0.258	0.339	0.116	0.274	0.355
		RF	0.410	0.187	0.231	0.404	0.202	0.245
	Alpha	**KNN**	0.115	0.258	0.337	0.093	0.280	0.361
		RF	0.343	0.194	0.238	0.298	0.212	0.256
P3	**Gamma**	**KNN**	0.078	0.267	0.344	0.100	0.280	0.359
		RF	0.285	0.199	0.243	0.401	0.201	0.245
	Beta	**KNN**	0.089	0.261	0.341	0.123	0.274	0.354
		RF	0.377	0.191	0.235	0.404	0.202	0.245
	Alpha	**KNN**	0.097	0.262	0.340	0.101	0.279	0.359
		RF	0.321	0.196	0.240	0.285	0.214	0.257
P4	**Gamma**	**KNN**	0.087	0.266	0.343	0.102	0.280	0.359
		RF	0.318	0.196	0.240	0.384	0.204	0.247
	Beta	**KNN**	0.112	0.259	0.338	0.120	0.274	0.355
		RF	0.398	0.188	0.232	0.366	0.206	0.249
	Alpha	**KNN**	0.098	0.263	0.341	0.079	0.284	0.363
		RF	0.322	0.196	0.240	0.291	0.213	0.256
PO3	**Gamma**	**KNN**	0.088	0.265	0.342	0.111	0.277	0.357
		RF	0.284	0.198	0.243	0.380	0.205	0.248
	Beta	**KNN**	0.128	0.255	0.335	0.125	0.272	0.354
		RF	0.397	0.189	0.233	0.374	0.206	0.248
	Alpha	**KNN**	0.107	0.262	0.339	0.093	0.281	0.361
		RF	0.260	0.201	0.245	0.243	0.217	0.260
PO4	**Gamma**	**KNN**	0.093	0.263	0.340	0.102	0.280	0.359
		RF	0.302	0.198	0.242	0.409	0.201	0.244
	Beta	**KNN**	0.095	0.262	0.341	0.103	0.277	0.358
		RF	0.413	0.187	0.231	0.407	0.202	0.245
	Alpha	**KNN**	0.096	0.263	0.340	0.112	0.278	0.357
		RF	0.270	0.200	0.244	0.251	0.216	0.260
P7	**Gamma**	**KNN**	0.096	0.264	0.342	0.106	0.278	0.357
		RF	0.291	0.198	0.242	0.398	0.202	0.246
	Beta	**KNN**	0.120	0.258	0.336	0.122	0.274	0.355
		RF	0.404	0.188	0.232	0.405	0.202	0.245
	Alpha	**KNN**	0.105	0.261	0.339	0.111	0.277	0.357
		RF	0.336	0.195	0.239	0.291	0.213	0.256
P8	**Gamma**	**KNN**	0.091	0.264	0.342	0.111	0.274	0.357
		RF	0.294	0.198	0.242	0.391	0.204	0.247
	Beta	**KNN**	0.119	0.258	0.337	0.116	0.274	0.356
		RF	0.399	0.188	0.232	0.401	0.202	0.245
	Alpha	**KNN**	0.109	0.261	0.339	0.091	0.281	0.362
		RF	0.329	0.195	0.239	0.285	0.213	0.257
T7	**Gamma**	**KNN**	0.106	0.261	0.338	0.118	0.273	0.355
		RF	0.310	0.197	0.241	0.417	0.200	0.243
	Beta	**KNN**	0.115	0.257	0.336	0.125	0.273	0.353
		RF	0.415	0.187	0.231	0.416	0.200	0.244
	Alpha	**KNN**	0.090	0.263	0.340	0.090	0.281	0.361
		RF	0.275	0.199	0.244	0.261	0.215	0.259
T8	**Gamma**	**KNN**	0.089	0.264	0.341	0.099	0.281	0.361
		RF	0.278	0.199	0.243	0.303	0.212	0.255
	Beta	**KNN**	0.095	0.261	0.340	0.107	0.277	0.358
		RF	0.330	0.195	0.239	0.324	0.210	0.253
	Alpha	**KNN**	0.099	0.263	0.341	0.091	0.281	0.361
		RF	0.248	0.201	0.246	0.250	0.216	0.260
O1	**Gamma**	**KNN**	0.072	0.268	0.344	0.107	0.278	0.357

		RF	0.322	0.195	0.240	0.378	0.205	0.248
	Beta	**KNN**	0.122	0.256	0.336	0.121	0.273	0.356
		RF	0.390	0.189	0.233	0.379	0.205	0.248
	Alpha	**KNN**	0.106	0.261	0.339	0.090	0.280	0.361
		RF	0.270	0.200	0.244	0.240	0.217	0.260
O2	**Gamma**	**KNN**	0.098	0.264	0.339	0.112	0.277	0.357
		RF	0.301	0.198	0.242	0.402	0.202	0.245
	Beta	**KNN**	0.091	0.263	0.342	0.095	0.279	0.360
		RF	0.396	0.189	0.233	0.404	0.201	0.245
	Alpha	**KNN**	0.102	0.262	0.339	0.099	0.279	0.359
		RF	0.284	0.198	0.243	0.248	0.216	0.260
Fz	**Gamma**	**KNN**	0.086	0.265	0.341	0.087	0.282	0.362
		RF	0.325	0.196	0.240	0.400	0.202	0.246
	Beta	**KNN**	0.097	0.262	0.340	0.114	0.275	0.355
		RF	0.419	0.186	0.230	0.413	0.201	0.244
	Alpha	**KNN**	0.104	0.260	0.338	0.103	0.279	0.358
		RF	0.265	0.200	0.245	0.231	0.217	0.261
Pz	**Gamma**	**KNN**	0.095	0.264	0.341	0.112	0.278	0.357
		RF	0.307	0.197	0.241	0.302	0.203	0.246
	Beta	**KNN**	0.108	0.261	0.338	0.120	0.274	0.355
		RF	0.412	0.187	0.231	0.387	0.204	0.247
	Alpha	**KNN**	0.083	0.264	0.342	0.092	0.282	0.361
		RF	0.318	0.195	0.240	0.273	0.215	0.258
Cz	**Gamma**	**KNN**	0.084	0.266	0.343	0.110	0.278	0.357
		RF	0.289	0.199	0.243	0.408	0.201	0.245
	Beta	**KNN**	0.105	0.260	0.339	0.107	0.278	0.358
		RF	0.400	0.188	0.232	0.390	0.204	0.247
	Alpha	**KNN**	0.095	0.262	0.340	0.087	0.282	0.361
		RF	0.342	0.195	0.238	0.300	0.212	0.256
Oz	**Gamma**	**KNN**	0.097	0.264	0.340	0.095	0.282	0.361
		RF	0.317	0.196	0.240	0.410	0.202	0.244
	Beta	**KNN**	0.095	0.263	0.341	0.112	0.276	0.356
		RF	0.418	0.186	0.230	0.395	0.204	0.246
	Alpha	**KNN**	0.097	0.262	0.341	0.092	0.280	0.361
		RF	0.315	0.196	0.240	0.281	0.214	0.257

Table C1- V/A estimation results obtained when comparing each electrode channel individually, using DEAP dataset.

Band	Feature	Estimator	Arousal			Valence		
			PCC	MAE	RMSE	PCC	MAE	RMSE
Gamma	H1	KNN	0.534	0.139	0.244	0.539	0.147	0.257
		RF	0.711	0.137	0.181	0.710	0.146	0.191
	H2	KNN	0.313	0.204	0.299	0.287	0.225	0.325
		RF	0.375	0.193	0.236	0.356	0.210	0.252
	H3	KNN	0.167	0.245	0.329	0.152	0.268	0.356
		RF	0.311	0.199	0.242	0.300	0.214	0.256
	SE	KNN	0.115	0.260	0.340	0.112	0.283	0.368
		RF	0.270	0.201	0.244	0.276	0.216	0.258
	WP	KNN	0.532	0.139	0.245	0.542	0.146	0.256
		RF	0.715	0.135	0.180	0.715	0.144	0.190
	WE	KNN	0.278	0.212	0.306	0.248	0.233	0.333
		RF	0.348	0.196	0.239	0.334	0.212	0.253
	IMFP	KNN	0.480	0.155	0.259	0.468	0.168	0.277
		RF	0.653	0.153	0.195	0.645	0.165	0.208
	IMFE	KNN	0.392	0.177	0.279	0.372	0.194	0.303
		RF	0.404	0.192	0.234	0.395	0.207	0.248
Beta	H1	KNN	0.444	0.165	0.267	0.446	0.176	0.282
		RF	0.637	0.154	0.197	0.630	0.167	0.210
	H2	KNN	0.430	0.169	0.271	0.410	0.184	0.292
		RF	0.460	0.185	0.227	0.442	0.201	0.242
	H3	KNN	0.278	0.210	0.305	0.261	0.230	0.328
		RF	0.376	0.193	0.236	0.348	0.211	0.252
	SE	KNN	0.162	0.247	0.331	0.146	0.268	0.355
		RF	0.326	0.197	0.240	0.306	0.214	0.255
	WP	KNN	0.463	0.160	0.262	0.456	0.173	0.280
		RF	0.654	0.150	0.194	0.643	0.162	0.207
	WE	KNN	0.381	0.180	0.281	0.361	0.199	0.306
		RF	0.410	0.189	0.233	0.387	0.208	0.249
	IMFP	KNN	0.343	0.198	0.293	0.318	0.210	0.314
		RF	0.416	0.214	0.241	0.421	0.213	0.240
	IMFE	KNN	0.330	0.203	0.304	0.299	0.220	0.323
		RF	0.402	0.176	0.217	0.423	0.189	0.220
Alpha	H1	KNN	0.302	0.206	0.299	0.258	0.229	0.326
		RF	0.508	0.176	0.219	0.459	0.196	0.238
	H2	KNN	0.365	0.186	0.286	0.325	0.206	0.312
		RF	0.411	0.191	0.233	0.374	0.208	0.250
	H3	KNN	0.232	0.226	0.316	0.183	0.252	0.346
		RF	0.354	0.195	0.238	0.310	0.213	0.255
	SE	KNN	0.137	0.251	0.336	0.120	0.275	0.361
		RF	0.313	0.198	0.241	0.280	0.215	0.257
	WP	KNN	0.321	0.202	0.295	0.268	0.227	0.324
		RF	0.514	0.174	0.218	0.467	0.194	0.237
	WE	KNN	0.376	0.182	0.283	0.342	0.203	0.308
		RF	0.433	0.186	0.229	0.396	0.194	0.247
	IMFP	KNN	0.299	0.229	0.300	0.275	0.229	0.310
		RF	0.457	0.185	0.213	0.423	0.187	0.215
	IMFP	KNN	0.276	0.230	0.312	0.274	0.230	0.314
		RF	0.409	0.191	0.217	0.390	0.192	0.220

Table C2- V/A estimation results obtained when comparing each feature extraction method individually, using DEAP dataset.